

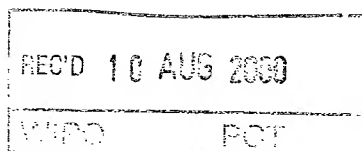


T/GB 00 / 02072



INVESTOR IN PEOPLE

GB00/2072



The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

4

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

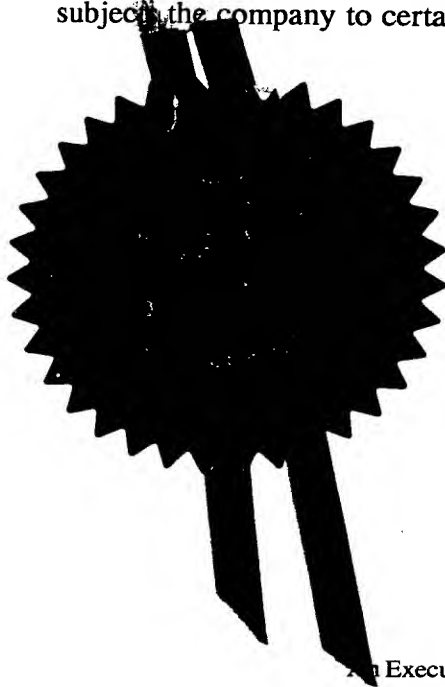
Re-registration under the Companies Act does not constitute a new legal entity but merely subject the company to certain additional company law rules.

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

Signed *Andrew Gersey*

Dated 14 July 2000



Patents 1977
16

The
Patent
Office

1/77

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

1. Your reference	IS/CP5714530		
2. Patent application number (The Patent Office will fill in this part)	28 MAY 1999		9912563.5
3. Full name, address and postcode of the or of each applicant (underline all surnames)	BIOTICA TECHNOLOGY LIMITED 112 HILLS ROAD CAMBRIDGE CB2 1PH		
Patents ADP number (if you know it)	7466324-001		
If the applicant is a corporate body, give the country/state of its incorporation	GB		
4. Title of the invention	POLYKETIDES AND THEIR SYNTHESIS		
5. Name of your agent (if you have one)	MEWBURN ELLIS		
"Address for service" in the United Kingdom to which all correspondence should be sent (including the postcode)	YORK HOUSE 23 KINGSWAY LONDON WC2B 6HP		
Patents ADP number (if you know it)	109006		
6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (if you know it) the or each application number	Country	Priority application number (if you know it)	Date of filing (day / month / year)
7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application	Number of earlier application		Date of filing (day / month / year)
8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer 'Yes' if: a) any applicant named in part 3 is not an inventor, or b) there is an inventor who is not named as an applicant, or c) any named applicant is a corporate body. See note (d))	YES		

Patents Form 1/77

9. Enter the number of sheets of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form 0

Description 147

Claim(s) 0

Abstract 0

Drawing(s) 4

10. If you are also filing any of the following, state

Priority documents 0

Translations of priority documents 0

Statement of inventorship and right 0

Request for preliminary examination 0

Request for substantive examination 0
(Patents Form 10/77)

Any other documents 0
(Please specify)

11. I/We request the grant of a patent on the basis of this application.

Signature

Date

27 May 1999

12. Name and daytime telephone number of person to contact in the United Kingdom IAN STUART

0117 9266411

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

- If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.
- Write your answers in capital letters using black ink or you may type them.
- If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
- If you have answered 'Yes' Patents Form 7/77 will need to be filed.
- Once you have filled in the form you must remember to sign and date it.
- For details of the fee and ways to pay please contact the Patent Office.

POLYKETIDES AND THEIR SYNTHESIS

The present invention relates to processes and materials (including enzyme systems, nucleic acids, vectors and cultures) for preparing polyketides, particularly polyethers but including polyenes, macrolides and other polyketides by recombinant synthesis, and to the polyketides so produced, particularly novel polyketides. (N.B the term "polyketide" is being used in its conventional sense to include structures notionally derived by the reduction and/or other processing or modification of one or more Ketide units). Furthermore the invention provides the entire nucleic acid sequence of the biosynthetic gene cluster that governs the production of the ionophoric antibiotic polyether polyketide monensin in *Streptomyces cinnamonensis*, and the use of all or part of the cloned DNA first, in the specific detection of other polyether biosynthetic gene clusters; secondly in the engineering of mutant strains of *S. cinnamonensis* and of other actinomycetes which are suitable host strains for the high level production of novel recombinant polyketides; and thirdly in the provision of recombinant biosynthetic genes which lead to such novel polyketide products.

Polyketides are a large and structurally diverse class of natural products that includes many compounds possessing antibiotic or other pharmacological properties, such as erythromycin, tetracyclines, rapamycin, avermectin, monensin, epothilones and FK506. In particular, polyketides are abundantly produced by

Streptomyces and related actinomycete bacteria. They are synthesised by the repeated stepwise condensation of acylthioesters in a manner analogous to that of fatty acid biosynthesis. The greater structural diversity found among natural polyketides arises from the selection of (usually) acetate or propionate as "starter" or "extender" units; and from the differing degree of processing of the β -keto group observed after each condensation. Examples of processing steps include reduction to β -hydroxyacyl-, reduction followed by dehydration to 2-enoyl-, and complete reduction to the saturated acylthioester. The stereochemical outcome of these processing steps is also specified for each cycle of chain extension. In addition, the biosynthetic pathways to many polyketides involve additional enzyme-catalysed modifications which may include: methylation by O- and C-methyltransferases, hydroxylation by cytochrome P450 enzymes, other oxidation or reduction processes, and the biosynthesis and attachment of novel sugars and/or deoxy sugars.

The biosynthesis of polyketides is initiated by a group of chain-forming enzymes known as polyketide synthases. Two classes of polyketide synthase (PKS) have been described in actinomycetes. One class, named Type I PKSs, represented by the PKSs for the macrolides erythromycin, oleandomycin, avermectin and rapamycin, consists of a different set or "module" of enzymes for each cycle of polyketide chain extension. (For examples see Cortés, J. *et al.* *Nature* (1990) 348:176-178; Donadio, S. *et al.* *Science* (1991) 252:675-679; Swan, D.G. *et al.*

Mol. Gen. Genet. (1994) 242:358-362; MacNeil, D.J. et al. Gene (1992) 115:119-125; Schwecke, T. et al. Proc. Natl. Acad. Sci. USA (1995) 92:7839-7843.)

The term "extension module" as used herein refers to the set of contiguous domains, from a β -ketoacyl-ACP synthase ("KS") domain to the next acyl carrier protein ("ACP") domain, which accomplishes one cycle of polyketide chain extension. The term "loading module" is used to refer to any group of contiguous domains which accomplishes the loading of the starter unit onto the PKS and thus renders it available to the KS domain of the first extension module. The length of polyketide formed has been altered, in the case of erythromycin biosynthesis, by specific relocation using genetic engineering of the enzymatic domain of the erythromycin-producing PKS that contains the chain releasing thioesterase/cyclase activity (Cortés J. et al. Science (1995) 268:1487-1489; Kao, C.M. et al. J. Am. Chem. Soc. (1995) 117:9105-9106).

In-frame deletion of the DNA encoding part of the ketoreductase domain in module 5 of the erythromycin-producing PKS (also known as 6-deoxyerythronolide B synthase, DEBS) has been shown to lead to the formation of erythromycin analogues 5,6-dideoxy-3- α -mycarosyl-5-oxoerythronolide B, 5,6-dideoxy-5-oxoerythronolide B and 5,6-dideoxy,6- β -epoxy-5-oxoerythronolide B (Donadio, S. et al. Science (1991) 252:675-679). Likewise, alteration of active site residues in the enoylreductase domain of module 4 in DEBS, by genetic engineering of the corresponding PKS-encoding DNA and its introduction into

Saccharopolyspora erythraea, led to the production of 6,7-anhydroerythromycin C (Donadio, S. et al. Proc. Natl. Acad. Sci. USA (1993) 90:7119-7123).

International Patent Application number WO 93/13663 describes additional types of genetic manipulation of the DEBS genes that are capable of producing altered polyketides. However many such attempts are reported to have been unproductive (Hutchinson, C.R. and Fujii, I. Annu. Rev. Microbiol. (1995) 49:201-238, at p. 231). The complete DNA sequence of the genes from *Streptomyces hygroscopicus* that encode the modular Type I PKS governing the biosynthesis of the macrocyclic immunosuppressant polyketide rapamycin has been disclosed (Schwecke, T. et al. (1995) Proc. Natl. Acad. Sci. USA 92:7839-7843). The DNA sequence is deposited in the EMBL/Genbank Database under the accession number X86780.

WO 98/01546 discloses that a PKS gene assembly (particularly of Type I) encodes a loading module which is followed by at least one extension module. The first open reading frame encodes the first multi-enzyme or cassette (DEBS1) which consists of three modules: the loading module (ery-load) and two extension modules (modules 1 and 2). The loading module comprises an acyltransferase and an acyl carrier protein. This may be contrasted with Figure 1 of WO 93/13663 (referred to above). This shows ORF1 as only two modules, the first of which is in fact both the loading module and the first extension module.

WO 98/01546 describes in general terms the production of a hybrid PKS gene assembly comprising a

loading module and at least one extension module. It also describes (see also Marsden, A.F.A. et al. Science (1998) 279:199-202) construction of a hybrid PKS gene assembly by grafting the wide-specificity loading module for the avermectin-producing polyketide synthase onto the first multi-enzyme component (DEBS1) for the erythromycin PKS in place of the normal loading module. Certain novel polyketides can be prepared using the hybrid PKS gene assembly, as described for example in WO 98/01571.

WO 98/01546 further describes the construction of a hybrid PKS gene assembly by grafting the loading module for the rapamycin-producing polyketide synthase onto the first multi-enzyme component (DEBS1) for the erythromycin PKS in place of the normal loading module. The loading module of the rapamycin PKS differs from the loading modules of DEBS and the avermectin PKS in that it comprises a CoA ligase domain, an enoylreductase ("ER") domain and an ACP, so that suitable organic acids including the natural starter unit 3,4-dihydroxycyclohexane carboxylic acid may be activated *in situ* on the PKS loading domain and, with or without reduction by the ER domain, transferred to the ACP for intramolecular loading of the KS of extension module 1 (Schwecke, T. et al. Proc. Natl. Acad. Sci. USA (1995) 92:7839-7843). WO 98/51695 and WO 98/49315 describe additional types of genetic manipulation of the DEBS genes that are capable of producing altered polyketides.

The second class of PKS, named Type II PKSs, is represented by the synthases for aromatic compounds. Type II PKSs contain only a single set of enzymatic activities

for chain extension and these are re-used as appropriate in successive cycles (Bibb, M.J. et al. EMBO J. (1989) 8:2727-2736; Sherman, D.H. et al. EMBO J. (1989) 8:2717-2725; Fernandez-Moreno, M.A. et al. J. Biol. Chem. (1992) 267:19278-19290). The "extender" units for the Type II PKSs are usually acetate units, and the presence of specific cyclases dictates the preferred pathway for cyclisation of the completed chain into an aromatic product (Hutchinson, C.R. and Fujii, I. Ann. Rev. Microbiol. (1995) 49:201-238). Hybrid polyketides have been obtained by the introduction of cloned Type II PKS gene-containing DNA into another strain containing a different Type II PKS gene cluster, for example by introduction of DNA derived from the gene cluster for actinorhodin, a blue-pigmented polyketide from *Streptomyces coelicolor*, into an anthraquinone polyketide-producing strain of *Streptomyces galileus* (Bartel, P.L. et al. J. Bacteriol. (1990) 172:4816-4826).

The minimal number of domains required for polyketide chain extension on a Type II PKS when expressed in a *Streptomyces coelicolor* host cell (the "minimal PKS") has been defined for example in WO 95/08548 as containing the following three polypeptides which are products of the *actI* genes: firstly KS; secondly a polypeptide termed the CLF with end-to-end amino acid sequence similarity to the KS but in which the essential active site residue of the KS, namely a cysteine residue, is substituted either by a glutamine residue or, in the case of the PKS for a spore pigment such as the *whiE* gene product (Davis, N.K. and Chater,

K.F. Mol. Microbiol. (1990) 4:1679-1691) by a glutamic acid residue; and finally an ACP. The CLF has been stated (for example in WO 95/08548) to be a factor that determines the chain length of the polyketide chain that is produced by the minimal PKS. However it has been found (Shen, B. et al. J. Am. Chem. Soc. (1995) 117:6811-6821) that when the CLF for the octaketide actinorhodin is used to replace the CLF for the decaketide tetracenomycin in host cells of *Streptomyces glaucescens*, the polyketide product is not found to be altered from a decaketide to an octaketide, so the exact role of the CLF remains unclear. An alternative nomenclature has been proposed in which KS is designated KS α and CLF is designated KS β , to reflect this lack of knowledge (Meurer, G. et al. Chemistry & Biology (1997) 4:433-443). The mechanism by which acetate starter units and acetate extender units are loaded onto the Type II PKS is not known, but it is speculated that the malonyl-CoA: ACP acyltransferase of the fatty acid synthase of the host cell can fulfil the same function for the Type II PKS (Revill, W.P. et al. J. Bacteriol. (1995) 177:3946-3952).

WO 95/08548 describes the replacement of actinorhodin PKS genes by heterologous DNA from other Type II PKS gene clusters, to obtain hybrid polyketides. It also describes the construction of a strain of *Streptomyces coelicolor* which substantially lacks the native gene cluster for actinorhodin, and the use in that strain of a plasmid vector pRM5 derived from the low-copy number vector SCP2* isolated from *Streptomyces coelicolor* (Bibb, M.J. and Hopwood, D.A. J. Gen. Microbiol. (1981)

126:427-442) and in which heterologous PKS-encoding DNA may be expressed under the control of the divergent *actI/actIII* promoter region of the actinorhodin gene cluster (Fernandez-Moreno, M.A. et al. J. Biol. Chem. (1992) 267:19278-19290). The plasmid pRM5 also contains DNA from the actinorhodin biosynthetic gene cluster encoding the gene for a specific activator protein, ActII-orf4. The ActII-orf4 protein is required for transcription of the genes placed under the control of the *actI/actIII* bidirectional promoter and activates gene expression during the transition from growth to stationary phase in the vegetative mycelium (Hallam, S.E. et al. Gene (1988) 74:305-320).

Type II clusters in *Streptomyces* are known to be activated by pathway-specific activator genes (Narva, K.E. and Feitelson, J.S. J. Bacteriol. (1990) 172:326-333; Stutzman-Engwall, K.J. et al. J. Bacteriol. (1992) 174:144-154; Fernandez-Moreno, M.A. et al. Cell (1991) 66:769-780; Takano, E. et al. Mol. Microbiol. (1992) 6:2797-2804; Gramajo, H.C. et al. Mol. Microbiol. (1993) 7:837-845). The DnrI gene product complements a mutation in the *actII-orf4* gene of *S. coelicolor*, implying that DnrI and ActII-orf4 proteins act on similar targets. A gene (*srmR*) has been described (EP 0 524 832 A2) that is located near the Type I PKS gene cluster for the macrolide polyketide spiramycin. This gene specifically activates the production of the macrolide antibiotic spiramycin, but no other examples have been found of such a gene. Also, no homologues of the ActII-orf4/DnrI/RedD

family of activators have been described that act on Type I PKS genes. WO 98/01546 describes the use of the ActII-orf4 family of activators in conjunction with their cognate promoters (e.g actII-orf4 with the *actI* promoter) in a heterologous actinomycete to obtain high level expression of recombinant Type I polyketide synthase genes.

Although large numbers of therapeutically important polyketides have been identified, there remains a need to obtain novel polyketides that have enhanced properties or possess completely novel bioactivity. The complex polyketides produced by Type I PKSs are particularly valuable, in that they include compounds with known utility as anthelmintics, insecticides, immunosuppressants, antifungal agents or antibacterial agents. Because of their structural complexity, such novel polyketides are not readily obtainable by total chemical synthesis, nor by chemical modifications of known polyketides.

There is also a need to develop reliable and specific ways of deploying individual genes and portions of genes in practice so that all, or a large fraction, of hybrid PKS genes that are constructed, are viable and produce the desired polyketide product. This includes the development of advantageous host strains for expression of such genes. For example many polyketides are rendered bioactive by the action of further enzymes other than the polyketide synthase, and host strains that contain and are able to express the genes for such enzymes are particularly convenient for the efficient synthesis of

the bioactive material. In those cases where the construction of a known or a novel polyketide requires specialised precursors, host strains containing and able to express the genes for key enzymes that enhance the production of such specialised precursors are equally valuable and desirable. There is also a need to develop rational methods of increasing the expression level of all the genes required for production of a specific polyketide. Clearly also a host cell which is advantageous for the above reasons, and/or because of other favourable characteristics including but not limited to its speed of growth, excellent handling characteristics in fermentation, and ease of transformation with DNA by various techniques, can be made even more favourable by the cloning into that cell of such auxiliary genes for polyketide modification, or gene activation, or post-translational modification, or precursor supply.

The DNA sequences have been disclosed for several Type I PKS gene clusters that govern the production of 16-membered macrolide polyketides, including the tylosin PKS from *Streptomyces fradiae* (application EP 0 791 655 A2), the niddamycin PKS from *Streptomyces caelestis* (Kavakas, S.J. et al. J. Bacteriol. (1997) 179:7515-7522) and the spiramycin PKS from *Streptomyces ambofaciens* (application EP 0791 655 A2). DNA sequences have also been disclosed for Type I PKS gene clusters that govern the production of further complex polyketides, for example rifamycin from *Amicolatopsis mediterranei* (WO

98/07868), and soraphen from *Sorangium cellulosum* (US 5716849), but so far no DNA sequence has been disclosed for one of the most widespread and important classes of complex polyketides, the polyethers.

5 Polyethers form an important group of complex polyketide antibiotics (Westley, J.W. in "Antibiotics IV. Biosynthesis" (Corcoran, J.W. Ed.), Springer-Verlag, New York (1981) p. 41-73). They are polyoxygenated carboxylic acids which act as selective ionophores transporting
10 cations across the cell membrane of target cells and thereby causing depolarisation and cell death. Certain polyethers including monensin, lasalocid and tetronasin are in widespread use in animal husbandry as coccidiostats (principally targetted against *Eimeria*
15 spp.) and as growth promoters. Polyethers have also been reported to be active *in vitro* and *in vivo* against the malarial parasite *Plasmodium falciparum* (Gumila, C. et al. Antimicrobial Agents and Chemotherapy (1997) 41: 523-529).

20 Polyethers contain multiple asymmetric centres and are characterised by the presence of tetrahydrofuran and tetrahydropyran rings, producing a characteristic shape which is non-polar on its outer surface and therefore well adapted for transport of material across bacterial
25 membranes; and provides on its inner surface polar coordinating ligands for a centrally-bound metal ion. In addition to tetrahydrofuran and tetrahydropyran rings, other groups which are often present include spiroketal, dispiroketal, and substituted benzoic acid moieties and
30 occasionally other groups for example a tetronic acid or

a 6-membered carbocyclic ring

Monensins A and B are produced by the actinomycete *Streptomyces cinnamonensis*. Their structures are shown in Figure 1. Monensin B differs from monensin A only in the presence of a methyl sidechain at C-16 rather than an ethyl sidechain. Monensin selectively binds and transports sodium ions. In addition to its antibacterial and antifungal properties monensin has some activity against protozoal parasites such as the malarial parasite *Plasmodium falciparum*. Although the structures of polyethers differ significantly from those of other complex polyketides such as the polyhydroxylated and polyene macrolides, their biosynthesis appears to take place by a metabolic pathway which has many common elements. Thus experiments using carbon 14-labelled precursors have shown that monensin A is synthesised from five acetate, one butyrate and seven propionate units (Day, L.E. et al. Antimicrob. Agents Chemother. (1973) 4:410-414). Similarly experiments using precursors doubly-labelled with carbon-13 and oxygen-18 have shown that oxygens (O)1, (O)3, (O)4, (O)5, (O)6 and (O)10 of monensin arise from the carboxylate oxygens of either propionate or acetate, while growth in the presence of oxygen-18 oxygen gas demonstrated that the three remaining ether oxygens (O)7, (O)8 and (O)9 are derived from molecular oxygen (Cane, D.E. et al., J. Am. Chem. Soc. (1981) 103:5962-5965; Cane, D.E. et al. J. Am. Chem. Soc. (1982) 104:7274 - 7281; Ajaz, A.A. and Robinson, J.A. J. Chem. Soc. Chem. Commun. (1983) 12:679-680). These findings have been rationalised by proposing that

the biosynthesis of monensin proceeds via an acyclic triene intermediate (1) in which the geometry of all three carbon-carbon double bonds is E (entgegen) rather than Z (zusammen). The triene is then proposed to be
5 subject to epoxidation to a tri-epoxide (2) and then ring opening is proposed to occur with concomitant sequential formation of the five ether rings as shown in Figure 2A. Such a biosynthetic pathway, first mooted by Westley in 1974 (Westley J.W. et al., J. Antibiot. (1974) 27:597-
10 604) accounts for the observed stereochemistry at the multiple asymmetric centres in monensin, (Cane, D.E. et al. J. Am. Chem. Soc. (1982) 104:7274-7281; Sood, G.R. et al. J. Chem. Soc. Chem. Commun. (1984) 21:1421-1424) and analogous schemes can be used to account for the
15 biosynthesis of other known polyethers. such as lasalocid A (Hutchinson C.R. et al., J. Am. Chem. Soc. (1981) 103:5953-5956), tetronasin (ICI 139603) (Demetriadou, A.K. et al. J. Chem. Soc. Chem. Commun. (1985) 7:408-410) and narasin (Spavold, Z. et al. Tetrahedron Letters
20 (1986) 27:3299-3302). The hydroxylation at C-26 and the introduction of an O-methyl group on oxygen 3 are proposed to occur as late steps in the biosynthesis, after formation of the polyether structure.

Unfortunately key aspects of the biosynthetic scheme
25 shown in Figure 2A have so far eluded experimental confirmation. No biosynthetic intermediates have been isolated from mutants of *S. cinnamonensis* that are blocked in early stages of monensin production. 26-deoxymonensin A has been isolated from a *S. cinnamonensis*
30 mutant partially blocked in monensin production

(Ashworth, D.M. et al. J. Antibiot. (1989) 42:1088-1099) and 3-0-demethylmonensins A and B have been recovered as minor components from the fermentation broth of a monensin-producing strain (Pospisil, S. et al. J. Antibiot. (1987) 40:555-557). When fed to cells of *S. cinnamonensis* in radio-labelled form, neither 26-deoxymonensin A, nor 3-0-demethylmonensin A, nor 3-0-demethyl, 26-deoxymonensin A were significantly incorporated into monensin A (Ashworth, D.M. et al. J. Antibiot. (1989) 42:1088-1099), either because they are actively excluded or because these modifications in fact occur earlier in the biosynthetic pathway so that these metabolites are shunt products not readily converted into the final antibiotic by the respective hydroxylase or methyltransferase. Similarly, the putative all (E)-triene precursor (1) has been synthesised and shown not to become incorporated into monensin when fed to growing cells of *S. cinnamonensis* (Holmes, D.S. et al. Helv. Chim. Acta (1990) 73:239-259). An alternative pathway has been proposed, as shown in Fig 2B, based on the transition-metal-mediated oxidation of 1,5-dienes (Walba, D.M. and Edwards, P.D. Tetrahedron Lett. (1980) 21:3531-3534). The triene intermediate (4) would differ from that of Figure 2A (1) only in that each carbon-carbon double bond would have the (Z)-configuration (Townsend, C.A. and Basak, A. Tetrahedron (1991) 47:2591-2602) and not the (E)- configuration.

The genetic basis of secondary metabolite biosynthesis essentially exists in the genes which code for the individual biosynthetic enzymes and in the

regulatory elements which control the expression of the biosynthetic genes. The genes encoding biosynthesis of polyketides in actinomycetes have hitherto been found as clusters of adjacent genes, ranging in size from
5 20 kilobasepairs (kbp) to over 100 kbp. The clusters often contain specific regulatory genes and genes conferring resistance of the producing strain to its own antibiotic.

10 In various of its aspects the invention provides the following:-

(1) a DNA sequence encoding at least one peptide necessary for the biosynthesis of monensin, preferably comprising one or more of the following genes: *mon BI*, *mon BII*, *mon CI*, *mon CII*, *mon H*, *mon RI*, *mon RII*, *mon T*,
15 *mon AIX* and *mon AX* as depicted in the appended sequence data or an allele or mutation thereof;

(2) a DNA sequence according to the first aspect comprising all of the genes listed therein or an allele or mutation thereof;

20 (3) a DNA sequence according to the first aspect comprising the complete monensin gene cluster;

(4) a DNA sequence coding for one or more of the peptides set out below, said peptide having the amino acid sequence as set out in the appended sequence data or
25 being a variant thereof having the specified activity:

<u>peptide</u>	<u>activity</u>
<i>mon CII</i>	epoxyhydrolase/cyclase
<i>mon E</i>	S-adenosylmethionine-dependent methyltransferase
<i>mon T</i>	monensin resistance gene
30 <i>mon RII</i>	repressor protein

	<i>mon AIX</i>	thioesterase
	<i>mon AI</i>	polyketide synthase multienzyme
	<i>mon AII</i>	polyketide synthase multienzyme
	<i>mon AIII</i>	polyketide synthase multienzyme
5	<i>mon AIV</i>	polyketide synthase multienzyme
	<i>mon AVI</i>	polyketide synthase multienzyme
	<i>mon AVII</i>	polyketide synthase multienzyme
	<i>mon AVIII</i>	polyketide synthase multienzyme
	<i>mon H</i>	regulatory protein
10	<i>mon CI</i>	flavin-dependent epoxidase
	<i>mon BII</i>	carbon-carbon double bond isomerase
	<i>mon BI</i>	carbon-carbon double bond isomerase
	<i>mon D</i>	cytochrome P450 hydroxylase
	<i>mon RI</i>	activator protein
15	<i>mon AX</i>	thioesterase

(5) a recombinant cloning or expression vector comprising a DNA sequence according to any of aspects 1-4;

(6) a transformant host cell which has been transformed to contain a DNA sequence according to any of aspects 1-4 and is capable of expressing a corresponding peptide;

(7) a hybridization probe comprising a polynucleotide which binds specifically to a region of the monensin gene cluster selected from *mon BI*, *mon BII*, *mon CI*, *mon CII*, *mon H*, *mon RI*, *mon RII*, *mon T*, *mon AIX* and *mon AX*;

(8) use of a probe according to aspect (7) in a method of detecting the presence of a gene cluster which governs the synthesis of a polyether, and optionally

isolating a gene cluster detected thereby;

(9) Use of a probe comprising a polynucleotide which binds specifically to a gene responsible for levels of activity of the monensin gene cluster, preferably a regulatory gene, resistance gene or thioesterase gene, more preferably the regulatory gene *mon RI*, in a method of detecting an analogous gene in a gene cluster of another polyketide, preferably a polyether, and optionally manipulating the gene detected thereby to alter the level of expression of said other polyketide;

(10) a host cell, preferably *Streptomyces cinnamonensis*, containing a heterologous gene under the control of the *mon RI* gene and a monensin promoter;

(11) use of a portion of the monensin gene cluster having chain terminating activity, preferably comprising at least one of *mon AIX* and *mon AX* or a mutant or allele thereof having chain terminating activity, to effect chain release of a peptide other than one required for monensin biosynthesis;

(12) use of a portion of the monensin gene cluster having carbon-carbon double bond isomerase activity, preferably comprising at least one of *mon BI* and *mon BII* or a mutant or allele thereof having isomerase activity to provide a desired stereochemical outcome in the synthesis of a polyketide other than monensin;

(13) a polypeptide encoded by a portion of the monensin gene cluster, preferably comprising at least one of *mon BI* and *mon BII* or a mutant or allele thereof, having carbon-carbon double bond isomerase activity;

(14) an epoxidase enzyme encoded by *mon CI* or a

derivative or variant thereof having epoxidase activity;

(15) a cyclase enzyme encoded by *mon CII* or a derivative or variant thereof having cyclase activity.

Some embodiments of the invention will now be described by way of example with reference to the accompanying drawings in which:

Fig 1 shows the structure of monensins A and B;

Fig 2 illustrates proposed biosynthetic pathways;

Fig 3 illustrates the proposed organization of the monensin polyketide synthase (PKS) enzyme complex; and

Fig 4 illustrates the proposed organization of the monensin biosynthetic gene cluster.

The overall gene organization of the monensin biosynthetic gene cluster, as shown in Fig 4, is similar to that previously found for many macrolide biosynthetic gene clusters, which have one or more open reading frames (ORFs) encoding large multifunctional PKSs flanked by other genes which encode functions required for the biosynthesis of the antibiotic. In the case of monensin, there is an unusually high number of distinct ORFs encoding PKS multi-enzymes (eight in total, labelled *monAI* to *monAVIII*) but there is again a separate module of enzymes for each cycle of polyketide chain extension, exactly as found for modular PKSs for macrolide biosynthesis (see Fig 3). Thus there are 12 condensations predicted to be required for the production of the carbon skeleton of monensin, and in agreement with this there are found to be 12 extension modules of PKS enzymes distributed among the 8 PKS ORFs. However, as mentioned in detail below, the other genes in the monensin cluster

include genes which have not previously been found in any other gene cluster for the biosynthesis of a complex polyketide, and which are not significantly similar to any genes in published sequence databases. The cloned DNA for these genes is useful to allow the diagnosis that a polyketide biosynthetic gene cluster in any actinomycete, uncovered previously by conventional hybridization against a PKS gene probe from (say) the DEBS or some other characterised PKS gene cluster, is one that governs the synthesis of a polyether; and these genes are also valuable either singly or in combination as specific hybridization probes for the specific detection and isolation of additional polyether biosynthetic gene clusters. Examples of these previously-unknown genes are the genes *monBI*, *monBII*, *monCI* and *monCII*. In addition the regulatory genes *monH* *monRI*, and *monRII* and the resistance gene *monT* and the thioesterase genes *monAIX* and *monAX* are all useful for the detection of analogous genes in other polyether clusters which are required for the rational manipulation of such genes in order to increase levels of the specific product.

The cloned and sequenced cluster of genes for monensin biosynthesis is useful secondly in the engineering of mutant strains of *S. cinnamonensis* and of other actinomycetes which are suitable strains for the high level production of either natural or novel recombinant polyketides. The sequence of the monensin cluster disclosed here shows the surprising fact, that the gene cluster contains a gene *monRI* whose gene product has an amino acid sequence highly similar to that of *actII*-

orf4, the pathway-specific activator gene which activates the *actI* and other promoters of the actinorhodin biosynthetic gene cluster of *Streptomyces coelicolor*. The recognition of this aspect of the natural regulation of a Type I PKS cluster is important and valuable because first, as will be immediately obvious to the person skilled in the art, it will be possible to increase the yield of monensin by increasing the level of the activator MonRI, either by placing the gene *monRI* under the control of a powerful promoter or arranging for the presence within the cells of one or more additional copies of the *monRI* gene; secondly, it will be possible to use the *monRI* gene as a specific hybridisation probe to locate similar genes in other complex PKS gene clusters, especially other polyether PKS gene clusters but also polyene and macrolide gene clusters and all other Type I modular PKS gene clusters; even in cases where (as for rapamycin and erythromycin) no such gene has been previously found within the currently accepted physical limits of the relevant biosynthetic gene cluster. In such cases the *monRI* gene probe might be expected to uncover the activator even if it resides on the chromosome at some distance from the main body of the gene cluster; and simple experiments would then show whether the activator(s) so uncovered are involved in regulation of the biosynthesis of those particular metabolites; thirdly, increasing the copy number of the *monRI* gene or of any of the activator genes uncovered will tend to increase the yield of a heterologous polyketide by "crosstalk" where the activator mimics the presence of the normal activator

for the transcription of the genes for that heterologous polyketide synthase. It is clear from recently published work (Wietzorrek, A. and Bibb, M. Mol. Microbiol. (1997) 25:1181-1184) that the ActII-orf4 family of activators
5 exert their effects by binding to promoter regions within the target gene cluster, so it will be possible to use the *monRI* gene together with monensin promoter regions to drive the high-level transcription and translation of heterologous genes in *Streptomyces cinnamomensis*, and
10 perhaps in other host strains too; such genes need not be PKS genes or even involved in polyketide biosynthesis. Monensin promoter regions are found at the 5' end of genes or groups of genes in the cluster and their location is clear from the sequence analysis disclosed here. Thus a
15 useful vector would provide the monensin promoter and the ribosome binding site and continue up to the start of the open reading frame, after which the monensin ORF naturally found there would be replaced by the heterologous gene. The relative strength of the monensin promoters can be
20 readily determined using any one of a number of known promoter probes, i.e. genes whose expression gives rise to readily measurable and quantifiable effects, such as Green Fluorescent Protein (GFP); or beta-galactosidase in the presence of a chromogenic substrate. It should be possible
25 to mutate randomly the small region of the monensin promoters especially likely to interact with the MonRI activator (identified by the presence of tandem heptanucleotide repeats with a common consensus sequence between the various monensin promoters) (Wietzorrek, A.
30 and Bibb, M. Mol. Microbiol. (1997) 25:1181-1184), and to

determine the optimal DNA sequence for the maximal
activation effect using either *S. cinnamonensis*
(preferably - in case there are other unknown factors that
make the activation function better in this strain than in
5 other heterologous systems), or even in another host
actinomycete strain. If the natural monensin promoters
were mutated to have this optimal recognition sequence,
then this would further increase the production of
monensin. By extension, the use of this modified monensin
10 promoter in conjunction with the *monRI* gene in
heterologous systems could form the basis of further
improvements in expression of polyketide synthases or
other genes, either by appropriate chromosomal alterations
to introduce the altered promoter and also the *monRI* gene;
15 or by provision of vectors containing these optimised
signals linked to specific genes and housed in suitable
host cells.

The sequencing of the monensin cluster has uncovered
another strategy for gene regulation in such Type I
20 clusters. The previously-sequenced genes for the rapamycin
biosynthetic pathway in *Streptomyces hygroscopicus*
included a gene of unknown function (*rapH*). A closely
similar gene has now been found in the monensin
biosynthetic gene cluster (*monH*), and it is clear from
25 this recurrence (and the comparison of the sequences with
those of database proteins) that this gene is potentially
an important DNA-binding sensor gene which acts to
regulate the transcription of the cluster in concert with
other regulatory signals. Simple experimentation is needed
30 in order to define whether the gene is an activator, in

which case putting in another copy or increasing its transcription will have the potential to increase polyketide biosynthesis; or alternatively the *rapH* gene product may be a negative regulator, whereupon deletion of this gene may release the biosynthetic pathway from this inhibitory effect and increase yields.

There is a continuing need to develop new methods of high-level production of bioactive metabolites and other valuable gene products in actinomycetes. *Streptomyces* *cinnamomensis* is a recognised and very valuable industrial strain for the production of very high levels of monensin, it is readily transformable with DNA by standard methods of conjugation or of protoplast transformation, it is a host for numerous known broad range plasmids including well-known expression plasmids of both high- and low-copy number, it also grows quickly relative to other actinomycete strains (for example about three times faster than wild type *Saccharopolyspora erythraea* the erythromycin producer, under comparable conditions) and sporulates relatively easily. Heterologous polyketides can be expressed in *Streptomyces cinnamomensis* using for example the low-copy number plasmid pCJR24 (which has no origin of replication active in actinomycetes so is maintained by integration into the chromosome) (Rowe, C. et al. Gene (1998) 216:215-223) or the related plasmid pCJR29 in which the polyketide synthase gene(s) are placed under the control of the *actI* promoter which is activated by the ActII-orf4 activator; or alternatively the *monAI* promoter can be substituted together with the MonRI activator; or some other pairing of activator and cognate

promoter chosen from either a Type II or a Type I polyketide synthase gene cluster. As an example, the wild type strain of *Streptomyces cinnamonensis* has been used to express the plasmid pCJR29 (Rowe, C. et al. Gene (1998) 216:215-223) containing as insert the three ORFs for the PKS governing the production of 6-deoxyerythronolide B, the macrolide precursor of erythromycin A in *Saccharopolyspora erythraea*, these genes being placed under the control of the pathway-specific *actI* promoter from *Streptomyces coelicolor* together with its cognate activator gene *actII-orf4*. The transformed strain when cultivated in a suitable liquid medium produced 6-deoxyerythronolide B in good yield.

It is well known to the person skilled in the art that it is possible to use standard vectors unable to replicate in actinomycetes to introduce DNA into a *Streptomyces* cell, such DNA comprising two portions of contiguous DNA which are each identical to one of two portions of the cell's chromosome that are spaced up to 100 kbp apart; and that through recombination between the incoming DNA and the chromosome occurring in both portions of DNA the net result is that the chromosomal sequence is replaced by the defective sequence originally that of the incoming DNA. Such a procedure has been applied to the monensin-producing strain of *S. cinnamonensis* as described in detail below, and a strain of *S. cinnamonensis* has been obtained that carries a specific deletion in the monensin cluster and which is unable to produce the antibiotic. The use of such a strain facilitates the production of heterologous polyketides by removal of the background of

monensin production.

The multiple uses of portions of the cloned and sequenced DNA from the monensin cluster will readily occur to the person skilled in the art. A surprising feature of the PKS of the monensin cluster is an unusual mechanism of polyketide chain initiation. We have found that the monensin PKS loading module has three domains, which from the amino-terminus of the protein are: a KSq domain, an acyltransferase domain and an ACP domain. We have uncovered this organisation in the PKS for the 14-membered macrolide oleandomycin as well as in the monensin PKS, an organisation of the loading module previously only found for the 16-membered macrolides and in which the KSq domain (which looks like a ketosynthase or condensation domain except that the active site cysteine residue is substituted by a glutamine for which the single letter notation is Q) had been previously speculated to have no function. It was realised that the acyltransferase of the loading module actually has malonyl-CoA and not acetyl-CoA as a substrate and that KSq is an active decarboxylase. It appears that a better discrimination can be achieved in the selection of the smaller acetate unit over propionate if the choice is made initially between methylmalonyl- and malonyl-CoA.

An unprecedented feature of the monensin PKS genes is that no integral chain-terminating domain is present as a C-terminal appendage of the PKS extension module that catalyzes the twelfth and final chain extension. Because the product of the monensin PKS is a carboxylic acid, it would have been firmly predicted that chain release would

have been catalyzed by such a C-terminal domain containing a "thioesterase" activity. Previously sequenced PKS gene sets have been of two sorts: first, those macrolide PKSs typified by erythromycin, spiramycin, tylosin, niddamycin
5 which have a readily recognisable C-terminal "thioesterase" domain, which in these enzymes functions as a specific cyclase rather than releasing the polyketide product as a free carboxylic acid; secondly, those macrolide PKSs typified by rapamycin, FK506, and
10 rifamycin, where there is an alternative and recognised mode of chain termination by transfer of the polyketide chain to an acceptor moiety, catalyzed by a specific enzyme (eg pipecolate incorporating enzyme for rapamycin (Schwecke T. et al. Proc. Natl. Acad. Sci. USA (1995)
15 92:7839-7843) and FK506 (Mothamedi H. and Shafiee A, Eur. J. Biochemistry (1998) 256:528-534); arylamine synthetase for rifamycin (August P.R. et al. Chemistry & Biology (1998) 5:69-79).

The monensin PKS surprisingly falls into neither
20 category, and therefore seems to be the first example of a novel mode of chain termination. It is novel and noteworthy in this connection that the monensin PKS gene cluster contains two small genes that encode discrete, monofunctional thioesterase enzymes. Although many PKS
25 gene clusters have been previously shown to contain one such discrete thioesterase, none have been shown to have two. The role of such thioesterases is not known, although in the case of methymycin/pikromycin PKS, which has been reported to be responsible for the biosynthesis of both
30 the 12-membered macrolide methymycin and the 14-membered

macrolide pikromycin (Xue Y.Q. Proc. Natl. Acad. Sci. USA
(1998) 95:12111-12116) the disruption of this thioesterase
reportedly caused a ten-fold drop in the amount of both
macrolides produced. A similar finding has been reported
5 for the discrete thioesterase of the tylosin PKS gene
cluster (Cundliffe E. *et al.* Chemistry & Biology in
press). Additional copies of such thioesterases may
therefore accelerate the production of specific
polyketide, but this has not yet been demonstrated.
10 However, the presence of the discrete thioesterase is not
completely essential for polyketide production.

It is highly desirable to have a broadly effective
method of catalysing the release of polyketide gene
products from a PKS as the free acid. The well-studied
15 integral thioesterase domain in the erythromycin PKS
thioesterase has a broad specificity in cyclization to
form a lactone (assuming that a hydroxy group is present
in the growing polyketide chain at an appropriate
position), but hydrolysis to form the free acid is very
20 slow. The recognition of the unusual arrangement of the
monensin PKS means that it is now possible to harness
either the entire PKS module that catalyses the twelfth
and final extension cycle in monensin biosynthesis, or the
C-terminal portion of it, and graft it onto a different
25 polyketide synthase by genetic engineering, so as to allow
the release mechanism characteristic of monensin to
operate in a different context. The use of this portion
only of the monensin PKS suffices to allow the novel
mechanism of chain release to operate successfully. The
30 speed of the polyketide chain hydrolysis in a given case

can depend on the additional presence of one or both of the discrete thioesterase genes (*monAIX* and *monAX*) from the monensin gene cluster. The use of this novel method of chain termination represents a valuable way of generating a large number of novel engineered polyketides that are currently inaccessible, and ensuring that the products have a specified chain length.

The genes *monBI* and *monBII* appear to encode very similar enzymes with significant amino acid sequence similarity to authentic ketosteroid isomerases which are known to catalyse the migration of an activated carbon-carbon double bond. The conservation of active site residues makes it very likely that these *mon* genes govern a reaction involving activated double bonds in the biosynthetic pathway to monensin and this surprising observation can be accommodated if the initial product of the polyketide chain growth on the monensin PKS is a linear precursor in which the double bonds were initially formed with a conventional *trans* or *E* (*entgegen*) geometry; but before the polyketide chain was extended by insertion of the next unit the *monBI* and/or the *monBII* gene product(s) catalyse the specific rearrangement of the newly-created double bond into the *cis* or *Z* (*zusammen*) geometry. This new view of the monensin biosynthetic pathway allows the deduction that the *monBI* and *monBII* genes, perhaps in combination with specific portions of the monensin modules where they normally exert their effects (namely modules 3, 5 and 7) might be used in order to achieve the extremely desirable targetted biosynthesis of novel polyketides containing double bonds with *Z*

geometry at specified point(s) along the chain. Thus for example it should be possible to provide for the direct biosynthesis of C22-C23 *cis* or *Z* double bond in avermectins, thus avoiding tedious and expensive chemical conversion of an initial fermentation product into this important anthelmintic. Only limited experimentation is needed to see whether the *monBI* and/or *monBII* gene products are sufficient or whether the *mon* PKS at modules 3, 5 and 7 forms part of the specific docking site(s) for the isomerases and therefore must also be used in the creation of the hybrid PKS that will insert the *cis* or *Z* double bond at the desired position. The substrate specificity of the isomerases need not be limited to 2,3-unsaturated thioesters. The purified enzymes could also be used to effect such isomerisations *in vitro*, depending on the position of the equilibrium or whether further enzymes are used to achieve the further transformation of the product as it is formed (*vide infra*).

The product of the *monCI* gene is a novel oxidative enzyme with some sequence similarity to authentic examples of such enzymes in the databases; and with a clearly definable role in the monensin biosynthetic pathway, the epoxidation of the double bonds at three separate positions in the initially-formed acyclic intermediate in monensin biosynthesis. This epoxidase could therefore be used in conjunction with *monBI/monBII* gene products to effect oxidative reactions on suitable substrates *in vitro* and *in vivo*. Similarly the *monCII* gene product is a putative cyclase that opens the epoxides and causes the formation of ether rings in monensin.

Any or all of the *monBI*, *monBII*, *monCI* or *monCII* genes may be introduced into a heterologous strain containing the gene cluster for another polyether, in order to divert the biosynthetic pathway and produce a polyketide of altered structure. In these experiments the analogues of these *monB* genes could either be present or (once located and characterised using the *mon* genes as probes) they may be deleted prior to the introduction of the *monB* and *monC* genes into that strain. The converse experiment in which analogues of the *monB* and *monC* genes from other strains are introduced into *S. cinnamonensis* likewise has the potential to produce novel oxidised polyketides. Also, the *monB* and *monC* genes or their analogues may be introduced into a strain that normally produces a macrolide or a polyene or some other complex polyketide and expressed there, when they may effect the diversion of the growing polyketide chain on a heterologous modular PKS towards a new product, which may or may not have the structure of a polyether.

The availability of the monensin gene sequence allows the institution of domain swaps to alter the acyltransferase (AT) specificity of a given module, for example the ethylmalonyl-CoA specific extender found in one of the modules of the monensin PKS can be used to replace one of the other ATs to generate an ethyl side branch at that position in the chain, or the AT can be used to substitute in any other (e.g. macrolide) PKS, as described in WO 98/01571 and WO 98/01546. Similarly the alteration of the level of reduction in a module, by

manipulation of the reductive enzymes, can be applied to the monensin genes and here it will produce, depending on which module is affected, either an altered monensin, or a species which is only partly cyclised, or a polyether with an altered pattern of cyclisation, or even a linear polyketide.

In general the targetted alteration of the pattern of substitution of sidechains or reduction level along the polyketide chain produced by the monensin PKS will, like the disruption or deletion of the oxidative enzymes mentioned above, lead to non-polyether polyketide products. It should be possible, by introduction of the DEBS thioesterase at the C-terminus of one of the later modules of the monensin PKS, together with an appropriately placed hydroxy group earlier in the chain, to produce novel macrolide products from this polyether PKS system, or alternatively novel polyenes of defined chain length and chosen ring size.

Example 1

Cloning of the monensin A biosynthetic gene cluster using
DNA probes derived from the erythromycin-producing
polyketide synthase of *Saccharopolyspora erythraea*

5 A genomic library of the monensin A producing strain
Streptomyces cinnamonensis ATCC 15413 was constructed
using methods well-known in the art, namely, the
production of high molecular weight genomic DNA, followed
by the partial cleavage of this DNA using the frequent-
10 cutting restriction enzyme *Sau*3A, fractionation of the
fragments on a sucrose gradient and selection of fragments
of average size 35-40 kbp, and the cloning of these
fragments into the cosmid vector pWE15 (Evans, G.A. et al.
Gene (1989) 79:9-20) which had been previously digested
15 with *Bam*HI and treated with shrimp alkaline phosphatase.
The library was packaged and transfected into *Escherichia*
coli XL-1 Blue MR cells. The library was plated out on
2xTY agar medium (10 g tryptone, 10 g yeast extract, 5 g
NaCl, 15 g bactoagar per litre containing ampicillin 50
20 μg/ml) for cosmid selection and the colonies were allowed
to grow overnight. The library was then screened by
hybridisation using as a probe DNA encoding the
ketosynthase domain of module 1 of the erythromycin-
producing PKS (6-deoxyerythronolide B synthase, DEBS) of
25 *Saccharopolyspora erythraea*. The colonies giving a
positive hybridisation signal in the hybridisation were
selected and the cosmid DNA from each colony was purified
and mapped by restriction digestion. The presence of the
target biosynthetic genes on a cosmid was verified by
30 sequencing of the ends of the cosmid inserts using the

commercially available T3 and T7 primers which hybridise specifically to the respective ends of each cosmid insert (Evans, G.A. et al. Gene (1989) 79:9-20).

Example 2

5 Sequencing of the biosynthetic gene cluster for monensin A
from *Streptomyces cinnamonensis*

Three cosmids obtained by screening of the genomic library of *S. cinnamonensis* were used to obtain the entire DNA sequence of the monensin biosynthetic gene cluster.

10 These cosmids, MO.CN02, MO.CN11 and MO.CN33 between them contain the entire DNA sequence of the cluster and the adjacent regions of the chromosome. They have been deposited in NCIMB, 23 St Machair Drive, Aberdeen AB24 3RY, UK, under the NCIMB accession numbers 40956
15 (MO-CN11); 40957 (MO-CN33) and 40958 (MO-CN02) respectively.

The DNA of each cosmid was separately subjected to partial digestion with *Sau3A* and fragments of approximately 1.5-2.0 kbp were separated by agarose gel
20 electrophoresis. The fragments were then ligated into the plasmid vector pUC18 (Messing, 1982), previously digested with *Bam*HI and treated with shrimp alkaline phosphatase. The library was transformed into *E. coli* strain XL1-Blue MR and plated on 2xTY agar medium containing ampicillin
25 (100 µg/ml) to select for plasmid-containing cells. Plasmid DNA was purified from individual colonies and sequenced using the Sanger dye-terminator procedure on an ABI 377 automated sequencer (Sanger, F. Science (1981) 214:1205-1210). The sequence data obtained from single
30 random subclones of a cosmid was assembled into a single

continuous sequence and edited using GAP4.1 program of the STADEN gene analysis package (Staden, R. Molecular Biotechnology (1996) 5:233-241).

The sequence is set out in the appended sequence listing.

Tables I and II contain data about individual genes and gene products.

Example 3

Inactivation of the monensin A biosynthetic gene cluster

A chromosomal gene disruption experiment was used to verify the identity of the cloned polyketide synthase gene cluster. Plasmid pMOB6314 is a pUC18 sequencing subclone of the presumed monensin A biosynthetic gene cluster prepared as described in Example 1, whose inserted DNA comprises the DNA sequence from nucleotide 9763 to nucleotide 10108 in SEQ ID 1, and which therefore contains a region of DNA wholly internal to *orfE*, a putative 3-O-methyltransferase. A *Hind*III fragment containing the thiostrepton resistance gene *tsr* from plasmid pIJ702 (Katz, E. et al. J. Gen. Microbiol. (1983) 129:2703-2714) was cloned into the *Hind*III site of plasmid pMOB6314 and the ligation mixture was used to transform *E. coli* cells. Transformants bearing the required plasmid pMOΔE01 were identified by isolation of plasmid DNA and analysis by restriction digestion. pMOΔE01. Plasmid pMOΔE01 was used to transform protoplasts of *Streptomyces cinnamonensis* as described by (Hopwood D.A. et al. (1985)). Since plasmid pMOΔE01 lacks an origin of replication that is active in *Streptomyces*, growth in the presence of thiostrepton (25 µg/ml) in the regeneration medium led to the isolation of

stable integrants. Isolated putative integrants were tested for the presence of integrated pMOΔE01 sequences by Southern hybridisation. A clone of *Streptomyces cinnamonensis* identified by its restriction pattern in Southern hybridisation as bearing pMOΔE01 integrated in the region of *monE* of the monensin A biosynthetic gene cluster was designated *S. cinnamonensis* MO-DD01.

Detection of production of the monensin A related metabolites produced by *S. cinnamonensis* MO-DD01 was performed by GC-MS analysis of methanol extracts of the entire broth harvested in 72 hours of growth of the strain. No significant amounts of monensin A-related metabolite production were detectable.

Example 4

Overproduction of erythromycin aglycone in *Streptomyces cinnamonensis*

S. cinnamonensis is a suitable system for overproduction not just of monensin A but also of other polyketide metabolites. Established techniques of genetic transformation allow fast introduction of foreign polyketide producing genes sets into this host. Fast growth of *S. cinnamonensis* in liquid culture and optimal precursor supply favour high yield of polyketide metabolites.

Protoplasts of *S. cinnamonensis* were prepared by a modified procedure of Hopwood et al. (1985). Plasmid pIB061 was transformed into the protoplasts of *S. cinnamonensis* and stable thiostrepton resistant colonies were isolated. Individual colonies were checked for their

plasmid content and the presence of plasmid pIB061 was confirmed by its restriction pattern. *S. cinnamonensis* (pIB061) was inoculated into 250 ml of M-C3 minimal production medium containing 10 μ g/ml of thiostrepton and allowed to grow for 72 hours at 30 °C. After this time the mycelia were removed by filtering. The broth was extracted with two volumes of ethyl acetate and the combined ethyl acetate extracts were washed with an equal volume of saturated sodium chloride, dried over anhydrous sodium sulphate, and the ethyl acetate was removed under reduced pressure to give about 200 mg of crude product. The product was analysed by LCQ and mass was confirmed to that of erythronolide B.

This example demonstrates the importance of *S. cinnamonensis* for production of high levels of foreign polyketide antibiotics. Introduction of the complete erythromycin gene cluster or other gene clusters into this system are likely to produce high levels of the corresponding metabolites.

References

1. Ajaz, A.A. and Robinson, J.A. (1983) The utilization of oxygen atoms from molecular oxygen during the biosynthesis of Monensin-A. *Journal of the Chemical Society-Chemical Communications*, **12**, 679-680.
2. Ashworth, D.M., Holmes, D.S., Robinson, J.A., Oikawa, H. and Cane, D.E. (1989) Selection of a specifically blocked mutant of *Streptomyces cinnamonensis* - isolation and synthesis of 26-deoxymonensin-A. *Journal of Antibiotics*, **42**, 1088-1099.
3. August, P.R., Tang, L., Yoon, Y.J., Ning, S., Muller, R., Yu, T.W., Taylor, M., Hoffmann, D., Kim, C.G., Zhang, X.H., Hutchinson, C.R. and Floss, H.G. (1998) Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. *Chemistry & Biology*, **5**, 69-79.
4. Bartel, P.L., Zhu, C.B., Lampel, J.S., Dosch, D.C., Connors, N.C., Strohl, W.R., Beale, J.M. and Floss, H.G. (1990) Biosynthesis of anthraquinones by interspecies cloning of actinorhodin biosynthesis genes in *Streptomyces* - clarification of actinorhodin gene functions. *Journal of Bacteriology*, **172**, 4816-4826.
5. Bibb, M.J., Biro, S., Motamedi, H., Collins, J.F. and Hutchinson, C.R. (1989) Analysis of the nucleotide sequence of the *Streptomyces glaucescens* *Tcml* genes provides key information about the enzymology of polyketide antibiotic biosynthesis. *EMBO Journal*, **8**, 2727-2736.
6. Bibb, M.J. and Hopwood, D.A. (1981) Genetic studies of the fertility plasmid SCP2 and its SCP2* variants in *Streptomyces coelicolor* A3(2). *Journal of General Microbiology*, **126**, 427-442.
7. Cane, D.E., Liang, T.C. and Hasler, H. (1981) Polyether biosynthesis - origin of the oxygen atoms of Monensin-A. *Journal of the American Chemical Society*, **103**, 5962-5965.
8. Cane, D.E., Liang, T.C. and Hasler, H. (1982) Polyether biosynthesis 2. Origin of the oxygen atoms of monensin A. *Journal of the American Chemical Society*, **104**, 7274-7281.
9. Cortés, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J. and Leadlay, P.F. (1990) An unusually large multifunctional polypeptide in the erythromycin producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature*, **348**, 176-178.
10. Cortés, J., Wiesmann, K.E.H., Roberts, G.A., Brown, M.J.B., Staunton, J. and Leadlay, P.F. (1995) Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science*, **268**, 1487-1489.
11. Davis, N.K. and Chater, K.F. (1990) Spore color in *Streptomyces coelicolor* A3(2) involves

the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. *Molecular Microbiology*, **4**, 1679-1691.

12. Day, L.E. (1973) *Antimicrobial Agents and Chemotherapy*, **4**, 410-414.
13. Demetriadou, A.K., Laue, E.D., Staunton, J., Ritchie, G.A.F., Davies, A. and Davies, A.B.
5 (1985) Biosynthesis of the polyketide polyether antibiotic ICI-139603 in *Streptomyces longisporoflavus* from O-18-labeled acetate and propionate. *Journal of the Chemical Society Chemical Communications*, **7**, 408-410.
14. Donadio, S., McAlpine, J.B., Sheldon, P.J., Jackson, M. and Katz, L. (1993) An erythromycin
10 analog produced by reprogramming of polyketide synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 7119-7123.
15. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J. and Katz, L. (1991) Modular
organization of genes required for complex polyketide biosynthesis. *Science*, **252**, 675-679.
16. Evans, G.A., Lewis, K. and Rothenberg, B.E. (1989) High efficiency vectors for cosmid
microcloning and genomic analysis. *Gene*, **79**, 9-20.
17. Fernandez-Moreno, M.A., Caballero, J.L., Hopwood, D.A. and Malpartida, F. (1991) The Act
15 cluster contains regulatory and antibiotic export genes, direct targets for translational control
by the *bldA* transfer-RNA gene of *Streptomyces*. *Cell*, **66**, 769-780.
18. Fernandez-Moreno, M.A., Martinez, E., Boto, L., Hopwood, D.A. and Malpartida, F. (1992)
Nucleotide sequence and deduced functions of a set of cotranscribed genes of *Streptomyces*
20 *coelicolor* A3(2) including the polyketide synthase for the antibiotic actinorhodin. *Journal of Biological Chemistry*, **267**, 19278-19290.
19. Gramajo, H.C., Takano, E. and Bibb, M.J. (1993) Stationary phase production of the antibiotic
actinorhodin in *Streptomyces coelicolor* A3(2) is transcriptionally regulated. *Molecular Microbiology*, **7**, 837-845.
20. Gumila, C., Ancelin, M.L., Delort, A.M., Jeminet, G. and Vial, H.J. (1997) Characterization
25 of the potent *in vitro* and *in vivo* antimalarial activities of ionophore compounds. *Antimicrobial Agents and Chemotherapy*, **41**, 523-529.
21. Hallam, S.E., Malpartida, F. and Hopwood, D.A. (1988) Nucleotide sequence, transcription
and deduced function of a gene involved in polyketide antibiotic synthesis in *Streptomyces*
30 *coelicolor*. *Gene*, **74**, 305-320.
22. Holmes, D.S., Sherringham, J.A., Dyer, U.C., Russell, S.T. and Robinson, J.A. (1990)
Synthesis of putative intermediates on the monensin biosynthetic pathway and incorporation
experiments with the monensin-producing organism. *Helvetica Chimica Acta*, **73**, 239-259.

23. Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M. and Schrempf, H. (1985) *Genetic manipulation of Streptomyces, a laboratory manual*. John Innes Institution, Norwich, UK.
24. Hutchinson, C.R. and Fujii, I. (1995) Polyketide synthase gene manipulation - a structure
5 function approach in engineering novel antibiotics. *Annual Review of Microbiology*, **49**, 201-238.
25. Hutchinson, C.R., Sherman, M.M., Vederas, J.C. and Nakashima, T.T. (1981) Biosynthesis of macrolides .5. Regiochemistry of the labeling of lasalocid a by C-13,O-18-labeled precursors. *Journal of the American Chemical Society*, **103**, 5953-5956.
- 10 26. Kakavas, S.J., Katz, L. and Stassi, D. (1997) Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*. *Journal of Bacteriology*, **179**, 7515-7522.
27. Kao, C.M., Luo, G.L., Katz, L., Cane, D.E. and Khosla, C. (1995) Manipulation of macrolide ring size by directed mutagenesis of a modular polyketide synthase. *Journal of the American*
15 *Chemical Society*, **117**, 9105-9106.
28. Katz, E., Thompson, C.J. and Hopwood, D.A. (1983) Cloning and expression of the tyrosinase gene from *Streptomyces antibioticus* in *Streptomyces lividans*. *Journal of General Microbiology*, **129**, 2703-2714.
29. MacNeil, D.J., Occi, J.L., Gewain, K.M., Macneil, T., Gibbons, P.H., Ruby, C.L. and Danis,
20 S.J. (1992) Complex organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase. *Gene*, **115**, 119-125.
30. Marsden, A.F.A., Wilkinson, B., Cortés, J., Dunster, N.J., Staunton, J. and Leadlay, P.F. (1998) Engineering broader specificity into an antibiotic-producing polyketide synthase. *Science*, **279**, 199-202.
- 25 31. Meurer, G., Gerlitz, M., Wendt Pienkowski, E., Vining, L.C., Rohr, J. and Hutchinson, C.R. (1997) Iterative type II polyketide synthases, cyclases and ketoreductases exhibit context-dependent behavior in the biosynthesis of linear and angular decapolyketides. *Chemistry & Biology*, **4**, 433-443.
32. Motamedi, H. and Shafiee, A. (1998) The biosynthetic gene cluster for the macrolactone ring
30 of the immunosuppressant FK506. *European Journal of Biochemistry*, **256**, 528-534.
33. Narva, K.E. and Feitelson, J.S. (1990) Nucleotide sequence and transcriptional analysis of the RedD locus of *Streptomyces coelicolor* A3(2). *Journal of Bacteriology*, **172**, 326-333.
34. Pospisil, S., Sedmera, P., Vokoun, J., Vanek, Z. and Budesinsky, M. (1987) 3-O-

Demethylmonensin-A and 3-O-demethylmonensin-B produced by *Streptomyces cinnamomensis*. *Journal of Antibiotics*, **40**, 555-557.

35. Revill, W.P., Bibb, M.J. and Hopwood, D.A. (1995) Purification of a malonyltransferase from *Streptomyces coelicolor* A3(2) and analysis of its genetic determinant. *Journal of Bacteriology*, **177**, 3946-3952.
36. Rowe, C.J., Cortés, J., Gaisser, S., Staunton, J. and Leadley, P.F. (1998) Construction of new vectors for high-level expression in actinomycetes. *Gene*, **216**, 215-223.
37. Sanger, F. (1981) Determination of nucleotide sequences in DNA. *Science*, **214**, 1205-1210.
38. Schwecke, T., Aparicio, J.F., Molnár, I., König, A., Khaw, L.E., Haydock, S.F., Oliynyk, M., Caffrey, P., Cortés, J., Lester, J.B., Böhm, G.A., Staunton, J. and Leadley, P.F. (1995) The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 7839-7843.
39. Shen, B., Summers, R.G., Wendtpienkowski, E. and Hutchinson, C.R. (1995) The *Streptomyces glaucescens* TcmK1 polyketide synthase and TcmN polyketide cyclase genes govern the size and shape of aromatic polyketides. *Journal of the American Chemical Society*, **117**, 6811-6821.
40. Sherman, D.H., Malpartida, F., Bibb, M.J., Kieser, H.M. and Hopwood, D.A. (1989) Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tu22. *EMBO Journal*, **8**, 2717-2725.
41. Sood, G.R., Robinson, J.A. and Ajaz, A.A. (1984) Biosynthesis of the polyether antibiotic Monensin-A - incorporation of [2-2-2H-2]-propionate, (R)-[2-2H-1]-propionate and (S)-[2-2H-1]-propionate. *Journal of the Chemical Society-Chemical Communications*, **21**, 1421-1423.
42. Spavold, Z., Robinson, J.A. and Turner, D.L. (1986) Biosynthesis of the polyether antibiotic narasin origins of the oxygen atoms and the mechanisms of ring formation. *Tetrahedron Letters*, **27**, 3299-3302.
43. Staden, R. (1996) The Staden sequence analysis package. *Molecular Biotechnology*, **5**, 233-241.
44. Stutzman-Engwall, K.J., Otten, S.L. and Hutchinson, C.R. (1992) Regulation of secondary metabolism in *Streptomyces* Spp and overproduction of daunorubicin in *Streptomyces peucetius*. *Journal of Bacteriology*, **174**, 144-154.
45. Swan, D.G., Rodriguez, A.M., Vilches, C., Mendez, C. and Salas, J.A. (1994) Characterization of a *Streptomyces antibioticus* gene encoding a type I polyketide synthase which has an

unusual coding sequence. *Molecular & General Genetics*, **242**, 358-362.

46. Takano, E., Gramajo, H.C., Strauch, E., Andres, N., White, J. and Bibb, M.J. (1992) Transcriptional regulation of the *redD* transcriptional activator gene accounts for growth phase-dependent production of the antibiotic undecylprodigiosin in *Streptomyces coelicolor* A3(2). *Molecular Microbiology*, **6**, 2797-2804.
47. Townsend, C.A. and Basak, A. (1991) Experiments and speculations on the role of oxidative cyclization chemistry in natural product biosynthesis. *Tetrahedron*, **47**, 2591-2602.
48. Walba, D.M. and Edwards, P.D. (1980) *Tetrahedron Letters*, **21**, 3531-3534.
49. Westley, J.W. (1974) *Journal of Antibiotics*, **27**, 597-604.
50. Westley, J.W. (1981) *Antibiotics IV. Biosynthesis*. Springer-Verlag, New York.
51. Wietzorrek, A. and Bibb, M. (1997) A novel family of proteins that regulates antibiotic production in *Streptomyces* appears to contain an OmpR-like DNA-binding fold. *Molecular Microbiology*, **25**, 1181-1184.
52. Xue, Y.Q., Zhao, L.S., Liu, H.W. and Sherman, D.H. (1998) A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: Architecture of metabolic diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 12111-12116.

1 GATCAGCGCG GTGGCGTCGT CGGCGTCCAG CTCGTTCTGC GTGGCGGACG
51 GCAGCGCGAT GTCGGCAGGC ACCTCCCAGA CCCGGCGGCC CGGCACGAAG
101 CGGGCCGAGG CGCCGCGGCG CTGGGCGTAG GTGTCCACGC GGGCGCGTTC
151 GACCTCCTTG ACCTGCTTGA GGAGGTCCAG GTCGATGCCC TTCTCGTCGA
201 CGACGTAACC GGAGGAGTCC GAACACGTCA CGGCGTTGGC GCCCAGGGCG
251 GCGAGCTTCT GGATGGTGTA GATGGCGACG TTCCCGGAGC CGGACACGAC
301 CGCCGTCCGG CCTTCGAGGG TCTCGCCGCG CTCACGCAGC ATCGCCGCCG
351 CGAAGAGGAC GTTGCCGTAG CCGGTCGCCT CCGGACGGAT CAGGGAGCCG
401 CCCAGTTGC GGCCCTTGCC GGTGAGGACG CCCGCCTCCC AGCGGTTGGT
451 GATGCGCCGG TACTGACCGA ACAGATAGCC GATCTCCCGG CCGCCGACGC
501 CGATGTCGCC CGCGGGCACG TCCGTGTGTT CGCCGATGTG CCGGTACAGC
551 TCCGTCAATGA ACGACTGGCA GAAACGCATG ACTTCCGCGT CGCTGCGGCC
601 GCGCGGGTCG AAGTCGCTGC CGCCCTTGCC GCCGCCGATG CCGAGGCCCCG
651 TCAGCGCGTT CTTGAAGATC TGCTCGAAGC CCAGGAACCTT GATGACGCCG
701 AGGTTACCG ACGGGTGGA GCGCAGGCCG CCCTTGTACG GGCCGAGGGC
751 GCTGTTGAAC TCCACCCGGA AGCCGCGGTT GACCCGCACG CGACCGTGGT
801 CGTCCTGCCA CGGCACCCGG AAGACGATCT GGCGCTCCGG TTCGCACAGG
851 CGCTCGATCA GGCCGGCTTC GGCCTACTCG GGGCGAGCCG CGATGACCGG
901 CGCCAGGGTC TCGAGGACCT CGCGGGCGGC CTGGTGGAA TCCGGCTGGG
951 CCGGGTTGCG GTGTTTCGATC TCGGTGAGCA GCTGGGAGAG TGCTGTCTTC
1001 TGCGAGAGAG CTGTCTTCGT GTCGGGTCGC GTGGTCAAAG GAGCCCTTTC
1051 TGGCACGGCC GGCCTAGGCG CTCGGCGCCG TTGCCGTGCG CAGGGAGACG
1101 CTCGAGCCGC AAGTATGACG CGCATGTAAA CACAGCGACC AGCCCCCGG
1151 TCCAGGGAGT GACCACCATG CGAGACCGGG CCACCGGTAG GGCCACCGGT
1201 CCGGCCTGCG GACCCCGTGT CACTTCCGGC TCGCGGCCAG GGGTGCCGCC
1251 CGGCGGACCG AATCGGCGGA GCGGCCAGC AGTGGCATGC GGACGGCCGG
1301 GCTGGGAATG CGGTTCTGGG CGTGCAGCAC TCCCTTGATC ACCGTCGGGT
1351 TCGGTTCCGGT GAAGAGGGCG GCGGAAAGGC GGGCGAGGTC GGCTCCGAGA

1401 GCGCGGGCGG GTGCGGCGGA GCCGCGTCGC CACAGCGCGA TCATCTCGGC
1451 GTAGTCGGCG GTACGCAGAT TGGCCGACGC CACGATTCCG CCGTGGGCGC
1501 CCGCAGCGAC CAGCGGCGAG AGGACGATGT CGTCACCGCC GAGCACGGCG
1551 AAGCCGGGCA GGGGCGAGTC GAGCAACTCC ATGGTGGTCG GGTCGATCGA
1601 GCCGGTCGCG TGCTTGATGC CGACGACCTC CGGCAGGCGG CCGAGTGC GG
1651 TGATCGTGCC CGCGCCGAGC GTCTGCCCCG TGCGGTAAGG GATGTCTGAC
1701 ACGACCAGGG GGAGGCCGCC GTGCTCGGCC AGCGCGGCGA AATGAGCCAG
1751 GGTCCCCGCT TCCCCGGGGC GGATGTAGGG CGGCGCGGGG ACCAGCGCGG
1801 CGGCGACGTC ACCCCGGGCC GCCAGCTCTC GCAGGGCCGT GATGGCGGTG
1851 GCGGTGTCGT TGGTGCCAC CCCGACGATG AGCGGTGCCC CGTGTGCCCC
1901 GCACGCGGCC GAGCAGACGC GGATCACCGT CTCTCTCTCC TCGGCGGTCA
1951 GTGTGGCGGC CTCGGCGGTC GTACCGAGGG CGACGAGCCC GGAGGCGCCG
2001 GCCGACAGCG CCTCGTCGGC GAGTCGGGCC AGCGCCTCGG GGGCCAGGCG
2051 CAGATCGTCG GTGAACGGAG TTACCAGGGG GACGTACAGG CCGTTGAAGA
2101 GCGGTTTCGGT GGTTCGGTTCG AGGCTCGATG CGAGGGTCAT GCTCTTACCC
2151 TGGCCACGC CACTCGGTAG ATCCATTTCA GATTCCTGCC GTCACACCTA
2201 AGCTGAACTT ATGCTCGATG TCCGTCGCCT CCATCTGCTC CGCGAACTCG
2251 ACCGGCGGGG CACCATCGCC GCCGTGGCCG AAGCGCTGAC CTTACCGCG
2301 TCCGCCGTCT CCCAGCAGCT CGGCGTGCTG GAGAGGGAGG CGGGCGTGCC
2351 GCTGTTGGAA CGCAGCGGCA GGC GCGTGGT CCTCACGCCC GCAGGACGCT
2401 CCCTCGTCGC ACACGCCGAC GCGGTGCTGA ACCGTCTCGA ACAGGCGGTC
2451 GCCGAGCTGG CGGGCGCACG GGACGGCATC GGCGGGCCGC TGCGCATCGG
2501 GACGTTCCCT TCCGGCGGCC ACACCATCGT CCCC GGCGCG CTGGCCGAAC
2551 TGGCCTCTCG TCACCCCGCG TTGGAGCCGA TGGTGCGGGA GATCGACTCC
2601 GCGCGCGTCT CCGACGGTCT GCGGGCCGGT GAGCTGGACG TGGCCCTCGT
2651 ACACGACTAC GACTTCGTAC CCGCGACGCC GGACACGACC GTGGACGAGG
2701 TGCCTCTGCT CGAAGAGCCG ATGTACCTCG TCACCCATGC CGCGGACACT
2751 GCCACGGACT CCGGCTCCGG GAGCACACTG GCAGCGCTGC TCGGGCCCTG

2801 TGCCGAGGTT CCGTGGATCA CGGCGCGGGA CGGCACGACC GGTCACGCGA
2851 TGGCTGTACG CGCCTGTCAG GCCGCCGGGT TCCAGCCCAG GATCCGCCAC
2901 CAGGTCAACG ACTTCCGCAC GGTGCTGGCT CTGGTCGCCG CCGGGCAGGG
2951 GGCCGGGTTC GTGCCGCGGA TGGCCGCCGA GCCGAGCCCC GCGGGCGTGG
3001 TGCTCACGAA GCTGCCGCTG TTCCGTCGCT CGAAGGTCGC GTTCCGTGCG
3051 GGCGGCGGTG CCCATCCGGC GATCGCCGCT TTCGTGGCCG CGGCGACGAC
3101 GGCGGTTCGAA CGCATGGCGG GTTCACGAGG CCCGGCCGGC GGCTCTGAGT
3151 GAACCGGCCG ACCGTGGGAA TGTGTGTGCC CTGGGCCGCA CCATTCTGTG
3201 CCTGGTGACG TCCTGGCGAC GTCCTGACGT CCTGATGTCC GAACGAGAAG
3251 GCGATTTTCC GCGATGGCCG ATGACGCGTA CCTGTTCTC CTCCCCGACC
3301 GGCACCCCCG ACTGGGAGCG GCCCTCGCCG CCGTCGGTGC CTTGGAATGC
3351 ACGGAAACCC CTGCGGTGCA CGCCTGGTTG CAGGCTCATG AGGCCTCCGT
3401 GTCCTCGGAA CAGGTCAGGA TTCTGCCCCG CGATGCCGAG AACTCATCC
3451 CGAAGGACGC CGAGCGGCTG CCGGTGCCGT TGAGCGAGGA GGAGGCGCTC
3501 AAGGTCGAGC AGGAGTGCGC GCCCCAGACC GTCACGGACA TGGAGAGCGA
3551 ACTGCTCGCG TTCCGGGAGA CGACCCAGGA CTGGCAGGCC CTCGTGCACC
3601 GGGCCCTGAC CGCGGGCATC CCCGCGCAGC GCATCGCCCG GCTGACCGGA
3651 CTCGACCCGG AGGAGATCGG CCGCCTGTAG GCGCTAGCGG CCGCCCAGTG
3701 CGGACACCAG GATGGCGACC GTGACGGTGT TGAAGACGAA GGCGATGACC
3751 GTGTTTCGCC CCACGGTCCG TCGCATGTCG CGTGAGGTGA CGTCGACATC
3801 GGTGGTGCCG AACGTCGTCA TCGCGGCCAG GGCGBAAATAG ACGTAGTCGG
3851 CCCAGGCGGG ACTCCGCTCC CCGGGGAATT CCAGTGCCCG CTCGTTCTCC
3901 ACGAGGTTGT CGGCCTGGAA GGTGACGGCG AAGGCCACGA CCACGCAGAT
3951 CCAGGCGGCG ACGACCAGGG CGAGCGCCAC CAGGGTGCGG GGGAGCGCGG
4001 AGAAGGTGGT GCTGAGGTGG CCGGGAAGCC ACAGCACCGC CACCACCAGC
4051 GCGGCAGCCG CGATGAAGAG CGAACCCCCG GGGCCGGGCG CGGTTCCGAG
4101 GACGTAACGC TGCAGGAATG TGCCGCGGGC TTCGCGCCGC GCCCAGGAGC
4151 GGACCTGCTC CGGAGCGACG CTCACGAAGA CGGTCATGGT GATGGCGAGG

4201 TAGGGCAGCA GGTAGGCGAA GAAGACGAGC ACGCCGACAT CCGCTGCCGA
4251 AATCCGCACC ACGGCGTCGA TGGGGAGGAC CACTGCCGCG CACGCCGCGA
4301 CGGCCAGGCT CACCGCCGAC CGGCGCCGTT CGGAAAGCCA GCGATGCACG
4351 GACGAGCCTC TCTGGTCGGG CGTCGGGCCT CGTGTGATCG TGACCGGCTC
4401 CGCGCCCGCC GAAAGCGCGG TGCATCTCC TGCCCTCGAA CGAGCGAAAC
4451 GCTTGCGCCG GAAAGCCTCC CTGCTGATGC CGACGGCGGC GGCAGTGGCT
4501 GCGGATGCGG ATCGTGCCTG GTGCCCTGAC CCTGGATGGG GGGAGGAACG
4551 CAGAGAGGCA GGTGCGCCCA TGACGGTCAT GGACAAGCTC AAGCAGATGC
4601 TCAAGGGGCA CGAGGACAAG GCCGGCCAGG GAATCGACAA GGCGGGCGAC
4651 TTCGTCGACG GGAAGACGCA GGGCAAGTAC AGCGGTCAAG TCGACACGGC
4701 CCAGGACAAG CTCCGGGACC AGTTCGGCTC GGATCAGCAG GAGCCTCCGC
4751 AGAGGTAGGC AGCGTCAGGG CGGAATCGGT CCGGGCGACC GCTGACCGCT
4801 GATGCAGATG CCGCAGACGT CGGCCCCGCA CTCCTCCGGG TAAATCGGAG
4851 CGTAGGCGGG GCCGACGTGT GCGCGTGCGG CCTCGTCTCT GCCGCCCCCTC
4901 TCCGCCCCGT CTCTGGCCCC TTGGTGCCAG TCTGACGGGA AAATGGCACC
4951 ACTTGGTGCC ACGCATGTGC CATGATGGCG TCATCGAGAG CGCGCTGCCC
5001 CGACTCGCGG GCAGGAAGGG CGCGTTCCGC GGAGTCGGCC GTCGGAGGGG
5051 TTGCATCATG GGGACAGCAC AGAGCCAGGA GCAGGCCGCC GCGCCCGGTG
5101 CCTGCGCCGC CTTCGTCCGC TTCGTGCTCT GCGGTGGCGG AGTGGGCCTC
5151 GCCTCCAGCT TCGCCGTGGT CGCCCTCGCC TCCTGGGTTC CCTGGGCGCT
5201 GGCCAACGCC CTGGTCGCCG TGGTCTCCAC CGTCGTCGCC ACCGAGCTCC
5251 ACGCCCGCTT CACCTTCGGT GCGGGCGGGC GCGCGACCTG GCGGCAGCAC
5301 GCGCAGTCGG CCGGGTCCGC GGCGGCCGCG TACGCGGTGA CCTGCGTGGC
5351 GATGTTTCGT CTGCAGCAGC TGGTGGCGGC GCGGCGCGCG GTGCTCGAGC
5401 AGGTGCTGTA CCTGTCGGCC TCCGCGCTCG CCGGTGTCGC GCGGTTCGTG
5451 GTGCTGCGCC TCGTCGTCTT CGCCCGGAAC CGCTCGCTGC CCGCCGCGGC
5501 CGCCGTGCGC ACCGCGCGTC CCGTGCGTCG CGTGCCGGCG CCCGTGCCCCG
5551 CGACCGTGGC CCACGCCGCA TCGCGCCCGG CCGGCCCCGC GCGCTCTGTC

5601 CCCGCCGCAT GACTCCGTGC CCGCATGTTT GTGCCCCCGG TGCTCCGTGC
5651 GTCCGGGGGC GGGGTGGGCG TCGTGCCCGG GTGGTCCAGG GGTCACGCGG
5701 TGGTGTGTGC CAGTTCCTGG CCGAGGTGGT GGGCGAGCTG TGCGGGCGTG
5751 GGGTTCTCGA CGATGGCGAC CATCGCGATC TCCATCCCGG TCAGCGTCAT
5801 CAGTGTCTTG GTGAGCTCAA GGGCGGTGAG GGAGTTGAGA CCGTTCTCGA
5851 GGAAGTTGCT GTCGTCGCTG AGGGTGGTGT TCAGAAGGGT GCCGGCCTGG
5901 GTGCGGATGG TGTCGGTGAG GAGCTTCTCG CGCTCCTCGG GGGTGGCCGC
5951 GGCGAGCTGC TTCTCCAGCT CGGTGGCGTC CTGGCCGGAG GTGTGGTCCG
6001 TGCTGGTCAT GACTGCTCCT GTGTGAGTGA GGTGTTGGCG GGGGTACAC
6051 CGCGGCGTGC GCGGTGTGGT CGTGCAGCCA GTAACGCGTG GCCTGGAAGG
6101 AGTACGTCGG GAGGTCGATG GTCCGGGGGT GGGGGGTGCG CCGGACGAGA
6151 GGGGTCCAGT CGACGGTGCC GCCCGTGGTG TGAAGCCGCG CGAGGGCGGT
6201 CAACAGGGCG CTTACGGCGG AGGTTTGCGT GCCTTCCGGT GAGAGCGCGC
6251 CCAGGTGGAG AAGCGTGTGG GTCTCGGGGG TGGGGGGTGC GGTGGGGGCC
6301 GGCGAGGTGA GGTGGTGGTG CCAGTAGTCG GCGGAGGCGA TGGGGGTGTC
6351 GGCCGGGGCA GTGCTGGTGA GCGTGAGCGT GGC GCGTTGG AACGTCAGCT
6401 GCTTCAGCAC GGGCTCGTAG GCGTCGGGCG GAGCCGGTTG TTCACCCTCG
6451 GCGGCCTGGG CGGCAGCGGC GTGGGCGGCG GCCAGGCGGC ACGCGTCGTC
6501 GAGGGTCAGG ATTCCCGCGG CGTACGCGGC GGC GATGTGG CCGACACCGT
6551 CGCCGGTGAG GGTGTGGGGG CGTACCCCCG TTTCCAGGAG CAGCCGCGCG
6601 AGCGCGGTGT GGACCGCGAA GCGCGCCAGT TCGGAGTGGG GAGTGGGGAG
6651 GGGAGTCGGC AGATGGGTGT CGAGGAGCGC GCGCGCTTCG TCGAAGGCGG
6701 ACGCGAAGAG CGGGAACGCC GAGTGGA ACT CGGCACCTCC GAAAGCCGCG
6751 CCGAATGTAG CGCCGAATGT CGCGCCGGGT TTGGCTCCGG GTGCCGCCCC
6801 CGTCGTCACC CCGTCGGCCG GGC GGCCGTC GAAGTGCCAG GCGATCTTCT
6851 TCGGGCCGGC CCCGGGCGTG GACCTGACCA GGTCCGGGTG GTCCTCTCCG
6901 GCGGCCAGGG CGCGGGCGGC GGC GAGGAGT TCGGTGTGGT CGGTGCCGGT
6951 GAGGACGGCG CGGTGTTC CA GGGGGCTGCG GGTGGCGGCG AGCGAGTAGG

7001 CGACCTCGGC GGGGGAGGGC GCGGGGTCGG TGGCCGCCAG GTGGGTGACG
7051 AGGGCCTTCG CCTGTGCCCC CAGGGCCTCG GGTGTACGAG CGGACAGGCT
7101 CCAGGCCACC GGGAGTTCCG GGGCAACGGG CGACGTCTGG TCGCGGGCGG
7151 CATCCGGCAC CGGAGCCTCG TCCACCGGCG GCTCTTCGAG GATGAGGTGC
7201 GCGTTCTGTC CGGACGTGGC GAAGGCGGAG ATGCCGACCC GGCGGGGCTC
7251 CTCGCGGCGG GGCCAGTCGA CCGCCTCGGT GAGCAGCCGT ACCGCGCCCT
7301 TCTTCCAGGC GGCGAGGGGC GTCGGGCGGT CGACGTGGAG GGTGCGCGGC
7351 AGGGTGCCGT GCCGGAACGC CTGGACCATC TTGATGAGCG CGGCCGCACC
7401 CGCGGCCCCC TGCGTGTGCC CCGTGTTGGA CTTGACGGAG CCGAGCCACA
7451 GGGGCCGGTC GGGGGAGCGG TCGGCGCCGT AGGTGGCGAG GAGGGCCTGG
7501 ACCTCGATGG CGTCGCCGAT GGGGGTGCCC GTCCCGTGCG CCTCGACGGC
7551 GTCGATCTGG TCCGGGGTGA GCCCGGCGTC GGCGAGGGCG GCGCGGATCA
7601 CATGCTGCTG GGAGGGGCCG TTGGGGGCGG CGAGGCCGTA TCCGGCGCCG
7651 TCCTGGTTGA CCGCGGAGCC GCGGATGACG GCGAGCACCG GGTGGCCGTT
7701 CTTCTGGCG TCGCCGAGCC GCTCAAGCAG GACGAGGCCG ACGCCTTCAC
7751 CGAGGCCCAT GCCGTCGGCC GCGGCGGCGA ACGGTTTGCA ACGGCCGTCC
7801 TGCGCGAGCG ACTTCTGGTG GGCGAAGGCG TGGAAGGTGT GCGGCGTCGA
7851 CATGACGGTG CCGCCGCCGG CGAGGGCGAG GCCGCACTCC CCGCGGCGCA
7901 GCGCCTGGCA GGCCAGGTGG AGGGCGACCA GGGAGGACGA GCAGGCCGTG
7951 TCCACGCTGA TGGCGGGGCC CTCGAGGCCG AGGGCGTAGG CGATGCGGCC
8001 GGAGACGAGG CTGCCGGACG TGCCGCCGCC CAGATAGGGC AGCAGCTCGT
8051 CGGGCGCGGT CTCGAGCCGT GTCGCGTAGT CGTGCCCGGT GGCGCCGACG
8101 TAGACGCCGG TGAGGGTGGA GCGCAGGGTG TGCGGGGCGA TGTGGCCGCG
8151 TTCGACGGTC TCCCACGCGA GGTGGAGCAT GAGGCGCTGG AGGGGTTCGG
8201 TGGCCACGGC CTCGGTGTCT CTGATGTCGA AGAAGCCCGC GTCGAAGCCG
8251 GCCGCGTCGT CCAGGAACCC GCCGAGCTCC GCGTACGGGC GTTCCTCGGG
8301 GAGTTCCCAG GCGCGGTCGT CGGGGAAGCC GGTGACGGCG TCGCGGCCCT
8351 CGGACACCAG ATCCCACAGG TCGTCCGGGG TGCGGGTCTT GCCGGGCAGC

8401 CGGCAGGCCA TGGAGACGAC GGC GATCGGC TCGTGCTGTG CGGCCTTCAG
8451 TTCGCGCAGC TGCTGCTGGG CCTGGTGGAG CTCGGCCGTC GTCCACTTGA
8501 GGTATTTCGAC GAGCTTCTCT TCGTTTCGCCA CGGGAATGGT CAGCCTTCCT
8551 GTTCTCGCGC GTGAAGCCTC AGGTGGGACG AGGTCGGGCA AGGTGGGCAG
8601 GCAGGAGCCG CGCGCTGTGG GTGCCAGGGT CGCCGCGGCT GCTTAAGCGG
8651 GTCTAACTCC CGCCTTGCCG CCGGGCATCG CCTCGCACGA GCGGGCCAGC
8701 AGCAGGAGGT CGGCGGCGAT CTCGTCTGGG GCGCCGGCGT GCAGATCGTG
8751 GTCGGAGCCC GGGTACCAGC GCACGCTCAC CTGCTCCAGG GCCGCCTCGG
8801 CGGCGGCCAC CCAGGCCCGT ACCTGGTCGG ACAGTTGGGG GATGGCGGGG
8851 ATGAGGGGCA GCAGCCGCAC CGGCACGGTG ACCTTGGGAT ACCAGTCGGC
8901 CGGTGCCTCC CGTTGCAGGC CGGCGACGAT CGACATGACC TGTGTCTGAGG
8951 TCAGGCGGGG GATGAGCAGG CCGTCCGGCC CGACGCGGTA GTCCGCCAGG
9001 CGTGCCTCGA TGGACGTGGG CGACCAGTCG GGATGGGTGG CCCGCAGGTA
9051 GGCGCGCATG TCGGCGGCGC TGGTGGTGCC CTGCTGGGCG CGCCGGACCA
9101 CGTCGGCGGT GCGCTCCCAG AAGGCGCGCA TCACCGGTCC GTCGAACTCG
9151 TACCAGCCGC CGTCGATCAG GGCGAGACCG GCCACCAGGT CCGGGTGCTC
9201 GGCCGCCAGG CGCAGCGCGA GGTGCGCGCC CCAGGAGTGC CCGGCCACCA
9251 GTGCGCCGGA CAGGTCGAGG GCGGTGACGG CCGCCACCAG GTCGGTGACG
9301 ACCGTCGCGT TGTCGTACCC GTCGGGCGGG GTGTCCGACT CGCCGTGGCC
9351 GCGGTGGTCG ACGGCGTAGG CCGGGTGTCC GGCGGCGGCG AGACGGGCGG
9401 CGACCTCGTC CCACATCCGG GCGTTCGACA GCATGCCGTG CAGCAGCAGG
9451 AACGGACGGC CCGGGGCTCC CGGCCCCTCC GCGGGCCGGT ACCTGACATT
9501 GAGGGAGACG GTCTGCGACA CGGGGATGCG GAGGTTCTTC ACAGGCGGGC
9551 CCTTGTGATC CCTTGTGCTG GGGGAGGAAA GCGGGGGCGG CACGCTCAGG
9601 GGCGCTGCGC GGTGCGGAAG ATGTATCCGA GCTCGGGCAT CTTGCCGAGG
9651 GCCGCCTGGT TGTGCAGGAA CAGCTCGTAT CCCTCTACGC CGATGATGTC
9701 GACGTACTCG TCCCGGTGGG CGCGGATCCA CTCGACGTAA CCGTCGTAGG
9751 TCTTGGCGGT CTCGCGGGTG ATGTCGGTCA GTTCGAGGAC GGTCCAGCCG

9801 GCGGCGCGGA AGATGTCGGG GTAGTCCCCG ATGTCGGTGA GCGGCGCGTA
9851 GATCGTGGTG TCGCTGACGG TGGCGGTCCG GGGCCGGCTG GGATCGGGGT
9901 TGAGGTAGAC CATGTCGGCG ATCGGCATCC GCGCGCCGGG CTTACGACG
9951 CGGTGGGCCT CGGTGAGCAC CTGCTGCTTG TCCGGCATGT GCAGCATGGA
10001 CTCCAGGGCC CAGCAGTGGT CGAAGGAGCC GTCGTCGAAC GGCAGGTTCA
10051 TGGCGTCGAC CTGCTCGAAG CGGACCCGGT CGGCGAGGCC GGCCTCGCGC
10101 GCGCGGCGGT TGCCGCGCTC GACCTGGCGG GCGCTGACGG AGATGCCGAC
10151 CACCTCGACG TCGCGGGCGC GGGCCAGCTG CATGGCCGGG GTGCCGTTGC
10201 CGCAGCCGAT GTCGAGGACG CGGTCGCCGG GGGCCGGGTC GAGGCGGCGG
10251 ATCATCTCGT CGGTCATCTG GACCATGGCC TCGTCGAACG TGGCCTGCTG
10301 CTCGCCGCCG TCGAACCAGT AGCCGTAGTG CAGATTGCCG TCTCCGAGCT
10351 GAGTCATCAG GTCGAAGACC TTGTGGTTCG AGTAGTGGCC GATGTCGCTG
10401 GGCTCGGGGG CGACGGTCTT GTTCACCGTC GGGGGCTTCT TGGTCGTCGC
10451 GTTCTTCGTC ACGGCTTCAG CGTCACCGTG CGGCGGCAGG CGCCACAACC
10501 CCACCCCCGC CCCTCAAAAG CCCCTATGGG CCCTCCTCGA CCGCCCCTAG
10551 GGAGCTGCTC TTGACGCGTT CCATACGGAA CGGGTGGTAC CCCTCCGAAA
10601 AAAATGAGAG TACGCTCCCA CTAGATATTG AGCTCTCTTT AGGAGGTCGA
10651 CTCCCATGTC TGCTGATCTG GGTGCGCGGC GGTGGTGGGC CGTCGGTGCT
10701 CTCGTA CTG CCGT CCGATGGT CGTGGGCTTC GATGTGACGA TCCTGAGCCT
10751 GGCGTTGCCC GCCATGGCCG ACGACCTCGG CGCGAACAAC GTCGAGCTGC
10801 AGTGGTTCGT GACGTCGTAC ACGCTGGTGT TCGCGGCCGG CATGATCCCCG
10851 GCCGGCATGC TCGGTGACCG GTTCGGACGC AAGAAGGTCC TGCTACCGC
10901 CCTGGTGATC TTCGGTATCG CCTCGCTGGC CTGTGCCTAC GCGACGTCCT
10951 CCGGCACCTT CATCGGCGCG CGTGCGGTGC TCGGTCTGGG CGCCGCGCTG
11001 ATCATGCCGA CGACGCTGTC GCTGCTGCCG GTCATGTTCT CCGACGAGGA
11051 GCGGCCGAAG GCCATCGGAG CGGTGGCCGG TCGCGCGATG CTCGCCTATC
11101 CGCTCGGCCC GATCCTCGGC GGCTACCTGC TCAACCACTT CTGGTGGGGC
11151 TCCGTCTTCC TGATCAACGT GCCGGTGGTG ATCCTCGCCT TCCTCGCGGT

11201 CTCCGCCTGG CTGCCCCGAGT CCAAGGCCAA GGAGGCCAAG CCGTTCGACA
11251 TCGGCGGCCT GGTGTTCTCC AGCGTCGGTC TCGCCGCGCT GACCTACGGC
11301 GTGATCCAGG GCGGCGAGAA GGGCTGGACG GACGTCACCA CGCTGGTGCC
11351 GTGCATCGGC GGTCTGCTCG CCCTCGTGCT GTTCGTGATG TGGGAGAAGC
11401 GGGTGGCGGA CCCGCTGGTC GACCTCTCGC TGTTCCGCTC GGCCCGGTTT
11451 ACCTCCGGCA CCATGCTCGG CACCGTCATC AACTTCACGA TGTTCCGGCGT
11501 GCTCTTCACG ATGCCGCAGT ACTACCAGGC GGTCCTCGGC ACCGACGCGA
11551 TGGGCAGCGG CTTCCGGCTG CTCCCGATGG TCGGCGGTCT GCTCGTGGGT
11601 GTGACGGTCG CCAACAAGGT CGCCAAGGCC CTCGGCCCCG AGACCGCGGT
11651 CGGCATCGGC TTCGCCCTCC TCGCCGCCGC CCTGTTCTAC GGCGCCACCA
11701 CGGACGTCAG CAGCGGCACC GGCCTGGCGG CCGCCTGGAC CGCGGCCTAC
11751 GGACTCGGCC TCGGCATCGC CCTGCCGACC GCCATGGACG CCGCCCTCGG
11801 CGCGCTCTCC GAGGACTCCG CCGGCGTCGG ATCCGGCGTC AACCAGTCCA
11851 TCCGTACCCT CGGCGGCAGC TTCGGCGCGG CCATCCTCGG TTCCATCCTC
11901 AACTCCGGCT ACCGCGGCAA GCTCGACCTC GACGGCGTGC CCGAGCAGGC
11951 ACACGGCGCG GTCAAGGACT CCGTCTTCGG CGGCCTCGCG GTGGCCCCGG
12001 CGATCAAGTC CAACGGACTG GCCGACTCGG TCGGTTCCGC GTACGTCCAC
12051 GCCCTGGACG TGGTGCTCGT GGTCTCCGGC GGCCTCGGAC TGCTGGGTGT
12101 GGTGCTGGCG GTGGTGTGGC TGCCCCGCCA TGTTGGTCAG AGCACCGCCA
12151 AGACAGCAGA ATCTGAGCAT GAAGCCGAG ACGCAGTCTG ACCAGGGCAA
12201 AACAGTGCCT GGTCTGAGAG AACGCAAGAA GGCCCGGACG AAGGCCGCGA
12251 TTCAGCGGGA GGCGGTGCGC TTGTTAGGG AACAGGGCTA CACCGCCACG
12301 ACCATCGAGC AGATCGCCGA AGCCGCCGAG GTCGCTCCCA GCACCGTCTT
12351 CCGCTACTTC GCGACCAAGC AGGACCTGGT CTTCTCGCAC GACTACGATC
12401 TGCCCTTCGC GATGATGGTC CAGGCCAGT CACCCGACCT GACGCCGATC
12451 CAGGCCGAGC GGCAGGCCAT CCGCTCGATG TTGCAGGACA TCAGCGAGCA
12501 GGAAGTGGCC CTGCAGCGCG AGCGGTTTCGT CCTGATTCTC TCCGAGCCGG
12551 AGCTCTGGGG CGCCAGCCTC GGCAACATCG GCCAGACCAT GCAGATCATG

12601 AGTGAGCAGG TGGCCAAACG GGCCGGGCGC GACCCGCGGG ACCCCGCGGT
12651 CCGCGCCTAC ACCGGAGCCG TGTTCGGAGT GATGCTCCAG GTCTCGATGG
12701 ACTGGGCCAA CGATCCGGAC ATGGACTTCG CGACCACGCT GGACGAGGCA
12751 CTCCACTACC TGGAAGACCT GCGGCCCTGA CCGAAGGGGC GGGCGCACAC
12801 CACAGAGCCC GCCCCGGCCA GACGTCGTAC GAGGCGCCAT CGGCCGTCGC
12851 GTACGACCCC CGCGCCCCGG ATTCCCCCGC GGGGCGCGGG GTCAAGGGAA
12901 AAGAGACGAC CGCACGCGGC CACTGTTCCC CCGGCTGCCG CGTCCGGTCC
12951 AACCTGGCGT GCTCCGGCTT CCCTCGACGG AGCACGCCAG GGGTCTGTCC
13001 GGCCCTCTCC CGGCGGCTCC CGTCAGACGC CCGGCCCCGC CGTCAGCGCC
13051 TCGGTCACGA CGGCCGCCAC CTGCTCCTGA CAGCCGTCGA GGTAAGAAGTG
13101 CCCGCCGGGC AGCACCCGCA GATCGAACGC CGCGCCGGTC CGTCCCCGCC
13151 ACGTGGCGGC CTGCTCCGGC GACGTCCGCT CGTCGGCGTC CCCGATCAGC
13201 GCCGTGATCG GGCAGTCGAG CCGGCCGGGT CCCGGCGCCT CGTAGGTGGC
13251 CACGGCCCCG TAGTCGGCCC GCAGCGCGGG CAGGACGAGC TCCTGCAGCT
13301 CGGGACTGCG GAAGAACCGC TCGTCGGTGC CGCCCATCGC CCGCAGATGG
13351 GCCAGGATGT CCGCGTCCCC GAACGCCCCC GAACGCCCCG CGGGACGGTA
13401 GGGCCGCGCG AGCCCCCGG AAACGAACAG GTGCACGGGA AGGCCGGGCC
13451 CGGCCGGTCC CCGCAGCCGC CGCGCCACCT CGAACGCCAC GATGGCGCCC
13501 AGGCTGTGCC CGAACAGCGC GAATGGCTTC CCGTCGCACG GCAGGTGGGG
13551 CACGACGCCG TCGGCGAGCT CGGCCACCGA CGCCAGGCAC GGCTCCGCAT
13601 GACGGTCCTG CCGCCCCGGA TACTGCACGG CGAGCACCTC GACGCCGGGC
13651 GCGAGCAGCC CGGAGAGCCC GAAGTAGTAA CTCGCCGAAC CGCCCGCGAA
13701 CGGAAAGCAG ACCAGCCGCA CCGGCGCCTC TGCCGCAGCG TGGTACCGCC
13751 GCAACCACAC CCCGTTTCCG GTGGCTGCAC CGAACTCGTC ACCGATCTGT
13801 GGTGCCCCGC CCGCCGTGCC CCTGTCCATC GTTCTCCCTC TCCTCGCGTC
13851 GCTCCGCGGG CGCTGTCTTG CCCC GCCCGG AAAGCCCGAT GCCGGCCAAG
13901 CCCC GATGCT GGCCAAACCC CGATGCCGGC CAAGCCCCGA TGCTGGCCGC
13951 GGCCCATAGC GCCCCGGCTAA AGCCGCAGGC GGCTAGCCGG GGT TTGGTTTC

14001 GCCTTTAGAC AGCCACCCA CGATGAGCCC GGTACTCGAA GCGATCTCCG
14051 ATTTCCGACC GGGAGCGCCG TTGATGTTTT GTGGCAGCCA GTTGTTTCAGC
14101 GCCCGACCGC AGCTGACGTG ATGGCCGCAT CCGCGTCAGC GTCCCCCTCG
14151 GGACCGAGCG CAGGACCCGA CCCGATCGCC GTCGTCGGGA TGGCCTGCCG
14201 CCTGCCCCGA GCACCTGACC CCGACGCGTT CTGGCGGCTG CTCAGCGAGG
14251 GGCGCAGCGC GGTGAGCACC GCACCGCCCG AGCGGCGGCG AGCCGACTCC
14301 GGCCTCCACG GGCCGGGCGG CTACCTGGAC CGGATCGACG GCTTCGACGC
14351 GGACTTCTTC CACATCAGCC CGCGCGAGGC CGTGGCGATG GACCCCCAGC
14401 AGCGGCTGCT CCTCGAACTG AGCTGGGAGG CCCTCGAAGA CGCGGGCATC
14451 CGGCCGCCCA CCCTGGCGCG CAGCCGCACC GCGTCTTCG TCGGCGCGTT
14501 CTGGGACGAC TACACCGACG TCCTGAACCT GCGGGCGCCG GGCGCCGTCA
14551 CCCGCCACAC CATGACCGGC GTGCACCGCA GCATTCTGGC CAACCGCATC
14601 TCGTACGCGT ACCACCTGGC CGGTCCGAGC CTCACCGTCG ACACCGCACA
14651 GTCCTCCTCG CTCGTCGCCG TCCACCTGGC CTGCGAGAGC ATCCGCAGCG
14701 GCGACTCCGA CATCGCCTTC GCGGGCGGCG TCAACCTCAT CTGCTCGCCG
14751 CGCACCACCG AGCTGGCCGC GGCCCGCTTC GGCGGTCTCT CGGCCGCAGG
14801 CCGCTGCCAC ACCTTCGACG CCCGCGCCGA CGGTTTCGTA CGCGGCGAGG
14851 GCGGCGGCCT CGTGGTGCTC AAGCCCCTCG CGGCGGCACG GCGCGACGGC
14901 GACACGGTGT ACTGCGTGAT CCGGGGGAGC GCCGTCAACA GCGACGGTAC
14951 GACCGACGGA ATCACCCTGC CCAGCGGGCA GGCGCAGCAG GACGTGGTGC
15001 GCCTCGCCTG CCGACGGGCG CGGATCACGC CGGACCAGGT GCAGTACGTC
15051 GAACTGCACG GCACCGGCAC GCCCGTCGGG GACCCGATCG AGGCCGCCCG
15101 GCTCGGCGCC GCCCTCGGGC AGGACGCCGC CCGCGCCGTG CCGCTGGCCG
15151 TCGGCTCCGC CAAGACGAAC GTCGGCCACC TCGAAGCCGC CGCCGGAATC
15201 GTCGGACTGC TCAAGACCGC CCTGAGCATC CACCACCGGC GGCTGGCGCC
15251 GAGCCTGAAC TTCACCACCC CCAATCCGGC CATCCCGCTC GCCGACCTCG
15301 GCCTGACCGT CCAGCAGGAC CTGGCCGACT GGCCGCGCCC CGAACAGCCC
15351 CTGATCGCCG GGGTGTGCTC CTTCCGCATG GGCGGCACGA ACGGTCACGT

15401 TGTCGTGGCG GCGGCGCCCG ATTCGGTGGC GGTACCTGAG CCGGTGGGGG
15451 TGCCTGAGCG GGTGGAAGTG CCTGAGCCGG TGGTGGTTTC TGAGCCGGTG
15501 GTGGTGCCGA CGCCATGGCC CGTGAGCGCT CACAGCGCTT CCGCGCTGCG
15551 CGCGCAGGCC GGTGCGCTGC GGACGCACCT CGCCGCCCAC CGCCCCACCC
15601 CCGACGCCGC GCGGGTCGGC CACGCGCTCG CCACCACCCG TGCGCCCCTC
15651 GCCCACCGCG CGGTCTTGCT CGGCGGCGAC ACCGCCGAAC TGCTGGGCTC
15701 CCTGGACGCG CTGGCCGAGG GCGCGGAGAC CGCGTCCATC GTGCGCGGCG
15751 AGGCGTACAC CGAGGGCAGG ACGGCCTTCC TCTTCAGTGG GCAGGGAGCG
15801 CAACGCCTCG GCATGGGGCG GGAGTTGTAT GCCGTGTTCC CCGTCTTCGC
15851 CGACGCTCTC GACGAGGCGT TCGCCGCCCT GGACGTACAT CTGGACCGCC
15901 CACTGCGCGA GATCGTCTTG GGCGAGACCG ACTCGGGTGG GAACGTCTCG
15951 GGTGAGAATG TCATCGGCGA GGGTGCCGAC CATCAGGCAC TCCTCGACCA
16001 GACCGCCTAC ACCCAGCCCG CGCTCTTCGC GATCGAGACG AGCCTGTACC
16051 GGCTGGCAGC CTCCTTCGGC CTGAAGCCGG ACTACGTCCT CGGCCACTCG
16101 GTCGGCGAGA TCGCCGCCGC GCACGTCGCC GGTGTCCTCT CGTTGCCGGA
16151 CGCGAGCGCT CTGGTGGCCA CGCGGGGACG GCTCATGCAG GCGGTTTCGG
16201 CGCCCGGCGC GATGGCCGCG TGGCAGGCCA CGGCGGACGA GGCGGCCGAA
16251 CAGCTCGCCG GGCACGAGCG GCACGTCACC GTGGCCGCCG TCAACGGCCC
16301 CGACTCCGTG GTCGTCTCCG GCGACCGCGC CACCGTCGAC GAACTGACCG
16351 CCGCCTGGCG GGGACGCGGC CGCAAGGCCC ACCACCTGAA GGTCAGCCAC
16401 GCCTTCCACT CCCCACACAT GGACCCCATC CTCGACGAGC TGCGCGCGGT
16451 CGCCGCCGGC CTGACCTTCC ACGAGCCGGT CATTCCCGTC GTCTCCAACG
16501 TCACCGGTGA ACTGGTGACC GCGACCGCGA CCGGGAGCGG CGCCGGGCAG
16551 GCCGACCCCG AGTACTGGGC GCGGCATGCG CGCGAGCCCG TGCGGTTCTC
16601 GTCCGGGGTG CGGGGGCTGT GCGAGCGCGG GGTGACCACG TTCGTCGAGC
16651 TCGGCCCGGA CGCACCGCTG TCCGCGATGG CCCGCGACTG CTTCCCCGCC
16701 CCCGCGGACC GGAGCCGTCC GCGCCCCGCC GCCATCGCCA CATGCCGCCG
16751 CGGGCGCGAC GAGGTGGCCA CGTTCCTGAG GTCGCTGGCC CAGGCGTACG

16801 TCCGCGGCGC CGATGTCGAC TTCACCCGGG CCTACGGCGC CACCGCCACG
16851 CGCCGCTTCC CCCTCCCCAC GTATCCCTTC CAGCGCGAGC GCCATTGGCC
16901 TGCCGCTGCC GGGGTGGGGC AGCAGCCGGA GACCCCGGAA CTTCCGGAAT
16951 CCTCGGAGTC CTCGGAGCAG GCAGGGCATG AGCGGGAGGA GGGGGCGCGC
17001 GCGTGGGGCG GGCCTGAAGG GCGGCTTGCC GGGCTCTCCG TGAACGACCA
17051 GGAGCGGGTC CTCCTCGGCC TGGTCACCAA GCACGTGGCC GTCGTGCTCG
17101 GGGACGCCTC GGGCACGGTA CAAGCCGCCC GCACCTTCAA GCAGTTGGGC
17151 TTCGACTCGA TGGCCGCCGC CGAGCTGAGC GAACGGGCTCG GCACGGAGAC
17201 GGGCCTGCCG TTGCCCGCCA CCCTCACCTT CGACTACCCG ACCCCTCTGG
17251 CCGTCGCCGC GCACCTGCGC GCGGAGCTCA CCGGTACGCC CGCCCCGGCC
17301 GGCTCCGCGC CCGCCACGGG CGCCCTCGGC GCGGGTGACC TCGGCACGGA
17351 CGAGGACCCG GTCGCCATCG TGGCCATGAG CTGCCGCTAT CCCGGCGGCG
17401 CAGGCACGCC CGAGGACCTG TGGCGGCTGG TCGCGGACGG CGCCGACGCG
17451 ATCGGAGACT TCCCCACCGA CCGCGGCTGG GACCTGGCGC GGCTGTTCCA
17501 CCCCAGCCCC GACCGGTCGG GCACCAGCTG CACGCGGCAG GGCGGATTCC
17551 TGTACGACGC CGCCGACTTC GACGCCGAGT TCTTCGACAT CAGCCCGCGC
17601 GAGGCCCTGG CCGTCGACCC GCAGCAGCGG CTGCTCCTCG AGTGCGCCTG
17651 GGAGGCCCTTC GAACGGGCGG GCCTGGACCC GCGGGCGCTC AAGGGCAGCC
17701 CCACCGGCGT GTTCGTCGGC ATGACGGGGC AGGACTACGG CCCCCGTCTG
17751 CACGAGCCGT CCCAGGCCAC CGACGGCTAT CTGCTGACCG GCAGCACGCC
17801 GAGCGTGGCC TCGGGCCGCC TGTGTTTCAG CTTGCGCCTT GAGGGGCCCCG
17851 CCCTGACGGT GGACACGGCC TGCTCGTCGT CGCTGGTCAC GCTCCATCTC
17901 GCGGCGCAGG CGCTGCGGCG CGGCAGGTGC GACCTGGCCC TCGCCGGCGG
17951 CGCCACCGTC CTGGCCACGC CGGGCATGTT CACCGAGTTC TCGCGGCAGC
18001 GGGGCCTGGC CCCCAGCGG CGCTGCAAGC CGTTCGCGGC GGGCGCCGAC
18051 GGCACGGGCT GGGCCGAGGG CGTGGGCCTG GTCCTCCTCG AAAGGCTCTC
18101 CGAGGCCCCG CGCAAGGGGC ACGCCGTCCT CGCGGTGATC CGGGGTTCGG
18151 CGATCAACCA GGACGGCGCG AGCAACGGCC TGACCGCGCC CAACGGCCCC

18201 TCGCAGCAAC GCGTCATCCG TGCCGCGCTC GCGGCCGCCC GGCTCACCGC
18251 GGACGAGGTC GACGTAGTGG AGGCGCACGG CACCGGCACC ACGCTCGGCG
18301 ACCCGATCGA GGCGCAGGCC CTGCTCGCCA CGTACGGCCA AGGGCGTTCTG
18351 GCGGAGCGGC CGTTGTGGCT CGGGTCGGTG AAGTCGAACA TCGGTCACAC
18401 GCAGGCCGCC GCGGGTGTCTG CGGGCGTCAT CAAGATGGTG ATGGCGATGC
18451 GCCACGACCT GCTCCCCGCC ACCCTGCACG TCGACGAGCC GAGTGGCCAC
18501 GTGGACTGGT CCACCGGCGC GGTGCGACTG CTCACCGAGC CGGTCGTCTG
18551 GCCGCGCGGC GAACGTCCGC GCCGCGCCGC GGTGTCGTCC TTCGGCATCT
18601 CCGGCACGAA CGCGCACCTG GTGCTCGAAG AGGCGGGGCA GGACGAGTAC
18651 GTTGCGGGAG CCGCCGACGA CGCCGGGCCG GTGGACGGTG CTGTGCTGCC
18701 GTGGGTGGTT TCCGGACGGA CCGGAGCGGC GCTGCGCGAA CAGGCCCGCC
18751 GTTTGCGTGA GTTGGTGACC GGC GGCTCGG CCGATGTCTC TGTGTCCGG
18801 GTGGGCCGGT CGCTGGTCAC CACGCGGGCG GTGTTCGAGC ACCGGGCCGT
18851 GGTCTGTGGC CGCGACCGGG ACACGCTGAT CGGCGGCCTC GAGGCCCTTG
18901 CGGCGGGTGA CGCGTCGCCG GACGTCGTGT GCGGGGTCGC GGGCGATGTC
18951 GGCCCCGGCC CGGTGCTGGT GTTCCCCGGG CAGGGCTCGC AGTGGGTGGG
19001 CATGGGAGCC CAACTCCTTG GCGAGTCCGC GGTGTTCGCG GCGCGGATCG
19051 ACGCGTGCGA GCAGGCGCTG TCCCCGTACG TCGACTGGTC ACTGACAGAG
19101 GTCCTGCGCG GGGACGGGCG CGAACTGTCG CGCGTCGACG TCGTCCAGCC
19151 CGTGCTGTGG GCGGTGATGG TCTCGCTCGC CGCCGTCTGG GCGGACCACG
19201 GCGTCACCCC GGCCGCCGTC GTCGGGCACT CCCAGGGAGA GATCGCCGCT
19251 GTGGTCGTCTG CCGGCGCGCT CACCCTGGAG GACGGCGCCA AGATCGTGGC
19301 CCTGCGCAGC CGGGCGCTGC GTCAGCTCTC GGGCGGGGGC GCCATGGCCT
19351 CCCTCGGGGT GGGCCAGGAA CAGGCAGCCG AACTCGTCGA GGGCCACCCC
19401 GGAGTGGGCA TCGCCGCCGT CAACGGCCCC TCATCGACCG TCATTTCAGG
19451 CCCGCCCCGAG CAAGTCGCCG CCGTCGTCTG CGACGCCGAG GCGCGCGAGC
19501 TGAGAGGCCG CGTCATTGAC GTGGACTACG CCTCGCACAG CCCCCAGGTC
19551 GACGCCATCA CCGACGAACT CACCCACACC CTGTCCGGCG TCCGCCCCAC

19601 CACGGCCCCG GTGGCGTTCT ACTCGGCCGT GACCGGAACC CGCATCGACA
19651 CGGCGGGCCT CGACACCGAC TACTGGGTCA CCAACCTGCG CCGCCCGGTC
19701 CGGTTGCGCG ACGCCGTCAC CGCGCTCCTC GCCGACGGCC ACCGGGTCTT
19751 CATCGAGGCC AGCAGCCACC CCGTCCTCAC CCTCGGCCTC CAGGAGACCT
19801 TCGAGGAGGC CGGGGTGCGAC GCCGTACCG TCCCCACCCT GCGGCGCGAG
19851 GACGGCGGCC GGGCACGCCT GGCCCGCTCG CTGGCACAGG CCTTCGGCGC
19901 CGGGTGCGCG GTGAGGTGGG AGAACTGGTT TCCGGCCACC GGTACGTCCA
19951 CCGTGGAGCT GCCGACGTAC GCCTTCCAGC GTCGCCGTTA CTGGCTGGAG
20001 GCCCCACGG GCACCCAGGA CGCGGCGGGC CTGGGCCTCG CCGCTGCGGG
20051 GCACCCGCTC CTCGGGGCGG CCACCGAGAT CGCGGACGGC GACATCCGCC
20101 TGCTCACCGG CCGTATCAGC AGGCACAGCC ACCCCTGGCT CGCTCAGCAC
20151 ACCCTCTTCG GTGCCGCGGT CGTGCCCGCC TCCGTCCTCG CGGAATGGGC
20201 GCTGCGCGCC GCCGACGAGG CCGGCTGCCC GCGTGTCGAC GACCTCACGC
20251 TGCGCACCCC GCTGGTGCTG CCCGAGACCG CGGGCGTGCA GGTGCAGATC
20301 GTGGTCGGCC CGGCCGACGC GCGGGACGGG CACCGCGACT TCCACGTCTA
20351 CGCCCCCCCC GACGGCAAGG ACGCCTCTGA GGGCGAGGGC ATCGCCGAGG
20401 GCGAGGGTGC CTCTGAGGGC GAGGGTGCCT CCGGCGGCAC CGATGCGCCC
20451 TGGACCTGCC ATGCCGACGG CCGACTGGTC GCCGAGCCCA CCGGCACGGC
20501 CTCGGAGGAC TCCCCGGACA CGGTGTGGCC GCCGCCCGGC GCCGAACCCG
20551 TCGACCTGGG CGACTTCTAC GAGCGGGCCG CCGCCACCGG AGTCGGCTAT
20601 GGACCGGTCT TCACGGGGCT GCGCGCCCTG TGGCGGCGGG ACGGCGAGCT
20651 GTTCGCCGAG GCGGTGCTGC CGCAAGAAGC CCCGAAACC GCCGGGTTCG
20701 GCATGCACCC GCGCTCCTC GACGCCGCAC TGCACCCCGC ACTCCTCGGC
20751 GAGCGGCCCG CCGAGGAGGA CAAGGTGTGG CTGCCGTTCA CGCTGACCGG
20801 AGTGACCCTG TGGGCCACCG GTGCCACCTC TGTACGCGTC CGTCTCACCC
20851 CGCTGGACGA CGACCCCGAC GCGTCGGCGG ACGGGCGGGC CTGGCGGGTC
20901 GGCGTGAGCG ACCCGACCGG CGCGGAGGTG CTGACCTGCG AGGCCCTGGT
20951 CGCGGTGGCG GCGGGCCGCC GCGAGCTGCG GGCCGCGGGG GAGCGGGTGT

21001 CCGATCTGTA CGCGGTGGAG TGGGTGCCGG TGCCGGGCCC GGGGCCGGTG
21051 GGTGAGGGTG CTGACTTCTC GGGCTGGGCC GGTCTGGGGG AGTGCGGGGA
21101 GCGTTGGGAG TCGTGGGGGC GCGTGGAGCG CTGGTACGAG GACCTGGACG
21151 CTCTCGGCGC GGCTGTCGAG GGTGGGGCTT CGGTGCCCTC TGTCGTTCTC
21201 GCCACCGCGG CTGCCGCCCC TGGTGGAGCG GGCACGGAG CCGCCGATGC
21251 GCTGAGCGCG GTGCGGTGGA CCGGCGCGCT CCTCGATCAG TGGCTCGCCG
21301 ACGCGCGGTT CGCCGACGCC CGGCTGGTGG TGATCACGTC CGGCGCGGTC
21351 GCCACGGGTG ACGATTTCCT TCCCGACCCG GCCGCCGCGG CGGTACGAGG
21401 ACTGGTCGAG CAGGCGCAGG TCAGGCACCC CGGCCGCATC CTCCTCGTCG
21451 ACACGGAAGC CGGGGCCGGG CTCGGGGTCG GCGCCGAGT GGATGACGCG
21501 CTCCTGGAAC AGGCCGTGGC CATGGCTCTC GGCGCCGACG AACCGCAACT
21551 CGCCCTGCGC GCGGGGCGGG TCCTGGCGCC CCGCCTCACC GCACCCAGG
21601 ATGCGGCCGT CACCGAAGCG GCGCGACCGC TCGACCCGGA CGGCACCGTA
21651 CTCATCACAG GGCCGGCCGG TGCTCCGGTG GCCGACCTCG CCGAACACCT
21701 CGTACGCACC GGGCAGTGCA GGCATCTGCT GCTCCTGCCT GGAGACGGTG
21751 AACTGGAGGA AATGGCCGAG GAGTTGCGGG GCCTCGGCGC CACCGTGGAC
21801 CTGAGTACCG CCGACCCGGC GGACCCGACC GCCCTCGCCG AAGTGGTCGC
21851 CGCCGTCGAG GGGGACCATC CTCTTACGGG GGTCATCCAC GCCACCGGAG
21901 TCGTGGACGC GTTCGATCCC GGCGACTCGG CGAGCGACTT GATGATCGAC
21951 TCGGCGAGCG ATTTCGTTGC CGAGGCATGG TCGTCGAGGG CGGGCGTCAC
22001 CGCCGCACTG CACACCGCGA CCGCCACCT TCCCCTGGAC CTGTTGCGCC
22051 TCCTGTCCCC GCGGGGCGCG GACCTGGGCA TTGCCCCGTC GCGGCCGCC
22101 GCGGGCGCCG ACGCCTTCAG CGCGGCACTC GCCCTGCGCC GGCACACGAC
22151 CGTCACGACG GACACGACAG CCCC GCCGCG CACGACAGCC CCGCCGCGAA
22201 CGACAGCCTC GCCGCGCACG ACAGCCCTGT CGTCGTCGCG CACGACGGGC
22251 GTGGCCCTCG CCTACGGGCC GCCCACC GCG CCGAGGCCCG GCATCAAGGG
22301 GACGGCGCCC GGTGCGATCC CCGTGCTGCT CGACGCCGCT CGCGCTCACG
22351 GGGGCGGTTT GCCCCTGCTC GGGGCCCGCT TGGCCGCGCG TGCCCTGGCC

22401 GCCGAGTCCG CCGCCGAGGG CGTCGCCGGC CTGCCC GCGC CGCTGCGCGC
22451 GCTGGCAGTG GCCGCAGCCG CGGCCGAGC ACCGACCCGG CGCACCGCCG
22501 CCGACCGCAA GCCCCCGCG GACTGGCCGG CCCGACTGGC CCCCCTGTCC
22551 GCCCCGAAC AACTCCGTCT GCTCATCGAC GCCGTACGA CCCACGCCGC
22601 CGCGGTCCTC GGCCGCACCG ACCCGGAAGC GCTGCGCGGG GACGCCACCT
22651 TCAAGCAGCT CGGCCTTGAC TCGCTGACCG CCGTGGAGCT GCGCAACCGG
22701 CTCGTGGAGG ACACCGGTCT GCGCCTGCCC ACCGCCCTCG TCTTTCGCTA
22751 CCCGACCCCC GCGGCGATCG CCGCGCACCT CCGCGAGCGG CTGACCAGCC
22801 CGAGCGAGAC GACCGCCACA CAGAGGTCCG GAGGGCAGAC GCCCGCAGCG
22851 GGGCAGGCGT CGTCCGCGCT CGCCCCCGGC GGATCGGCCG CCGGACCGCC
22901 CGCCGCAGAC ACCGTGCTGA GCGACCTGAC CCGCATGGAG AACACCCTCT
22951 CCGTGCTCGC CGCCCAGCTG CCCCACACCG AGACGGGTGA GATCACCACC
23001 CGGCTCGAAG CGCTCCTCAC GCGCTGGAAG ACCACGAACG CCACGGCGAA
23051 CGACAGCGGC GACGGCAACG GCGGCGATGA CGACGCCGCC GAACGCCTCA
23101 AGGCCGCGTC CGCCGACCAG ATCTTCGACT TCATCGACAA CGAGCTTGGT
23151 GTCGGGCACG GCACCTCGCG CGTGACCCCC ACTCCGAAGG CCGGGTGACC
23201 GCACATGGCG AGTGAAGAGC AACTGGTCGA ATATCTGCGC AGGGTGACCA
23251 CCGAGCTCCA TGACACGCGT CGGCGCCTGG TGCAGGAGGA GGACCGCAGG
23301 CAGGAACCGG TGGCCCTGGT CGGCATGGCC TGCCGCTTCC CGGGCGGCGT
23351 GGCCTCACCG GAGGACCTCT GGGACCTGGT CGCCGCGGGC AAGGACGCCA
23401 TCGAGGACTT TCCCACCGAC CGGGGCTGGG ACCTGGAGGC GCTCTACGAC
23451 CCGGACCCGG CCGCGTACGG GACCAGCTAT GTCCGCCACG GCGGGTTCGT
23501 GGACGACGCG GGCTCCTTCG ACGCCGACTT CTTGCGCATC AGCCCGCGAG
23551 AAGCCCTGGC GATGGACCCG CAGCAGCGGC TGATGCTGGA GACGTCCTGG
23601 GAGCTGTTCG AGCGCGCCGG CATCGAACCC GTCTCCCTCA AGGGCAGCCG
23651 TACGGGCGTC TACGCCGGGG TGTCCAGCGA GGACTACATG TCCCAACTGC
23701 CCCGCATCCC CGAGGGGTTC GAGGGGCACG CCACCACCGG CAGCCTCACC
23751 AGCGTCATCT CGGGCCGGGT CGCGTACAAC TACGGCCTCG AAGGCCCGGC

23801 CGTCACCGTC GACACAGCCT GTTCCGCCTC GCTCGTCGCC ATCCACCTGG
23851 CGAGCCAGGC GCTGCGCCAG CGTGAGTGCG ACCTCGCCCT CGCGGGCGGT
23901 GTGCTCGTAC TGTCCAGCCC GCTCATGTTC ACCGAGTTCT GCCGCCAGCG
23951 GGGCCTTGCT CCCGACGGCC GCTGCAAGCC GTTCGCCGCC GCGGCGGACG
24001 GCACCGGCTT CTCGGAGGGC ATCGGTCTGC TCCTCCTGGA GCGCCTGTCC
24051 GACGCGCGCC GCAACGGCCA CAAGGTGCTC GCGGTGATCC GCGGCTCCGC
24101 CGTCAACCAG GACGGCGCGA GCAACGGCCT GACCGCCCCC AACGACGCCG
24151 CGCAGGAACA GGTCATCCGC GCCGCCCTCG ACAACGCCCCG CCTCACCCCG
24201 TCCGAGGTGG ACGCCGTCGA GGCGCACGGC ACCGGCACCA AACTGGGCGA
24251 CCCCATCGAG GCCGGAGCGC TGCTCGCCAC CTACGGGCAA CACCGCGCCC
24301 GGCCCTCCTT CCTCGGCTCC CTCAAGTCCA ACATCGGCCA CACCCACGCC
24351 ACCGCGGGCG TCGCCGGTGT CATCAAGACC GTCATGGCGA TCCGCAACGG
24401 TCTGCTCCCC GCCACCCTCC ACGTCGAGGA ACTGAGCCCG CACGTCGACT
24451 GGGACGCGGG CGCGGTTCGAG GTCGTCACGG AGCCCACCCC GTGGCCCGAG
24501 ACCGGCCACC CCCGGCGCGC GGGCGTCTCC GCGTTCGGGA TCTCCGGGAC
24551 GAATGCGCAC TTGATCCTGG AGGAGGCGCC GCCGGAGGAG GATGTGCCCCG
24601 CCCCCGTGGT TGTGGAGTCG GGCGGGGTCG TTCCGTGGGT GGTGTCCGGG
24651 CGGACGCCCG AGGCGCTGCG TGAACAGGCC CGGCGACTCG GCGAGTTCGT
24701 GGCAGGCGAC ACGGACGCAC TGCCGAACGA GGTCGGCTGG TCCTTGGCCA
24751 CGACCCGGTC GGTGTTTCGAG CACCGGGCTG TGGTCGTGGG GCGTGACCGG
24801 GATGCGTTGA CGGCTGGCCT GGGGGCGTTG GCTGCGGGTG AGGCTTCGGC
24851 GGGTGTGGTG GCCGGGGTGG CCGGTGATGT GGGTCCTGGG CCGGTGTTGG
24901 TGTTTCCGGG GCAGGGGGCG CAGTGGGTGG GCATGGGTGC CCAGCTGTTG
24951 GACGAGTCTG CGGTGTTTCG GCGCGGGATC GCGGAGTGTG AGCGGGCCCT
25001 GTCGGCGCAT GTGGAAGTGT CGCTGAGTGC GGTGTTGCGC GGGGACGGGA
25051 GTGAGCTGTC CCGGGTGGA GTGGTGACGC CGGTGCTGTG GGCGGTGATG
25101 GTCTCGCTGG CTGCGGTGTG GGCGGATTAC GGGGTCACTC CCGGTGCCGT
25151 GATCGGGCAC TCGCAGGGTG AGATGGCTGC CGCGTGTGTG GCGGGGGCGC

25201 TGTCGCTGGA GGATGCGGCG CGGATCGTAG CGGTACGCAG TGACGCGCTT
25251 CGTCAGCTGC AAGGGCACGG CGACATGGCC TCGCTCAGCA CCGGTGCCGA
25301 GCAGGCCGCT GAGCTGATCG GTGACCGGCC GGGCGTGGTC GTCGCGGCGG
25351 TCAATGGGCC GTCGTCTACG GTGATTTTCTAG GGCCGCCGGA GCATGTGGCA
25401 GCCGTGGTCG CGGATGCGGA GGCACGTGGT CTGCGCGCCC GTGTCATCGA
25451 CGTCGGCTAT GCCTCGCATG GCCCCCAGAT CGACCAGCTC CACGATCTGC
25501 TGACCGAACG CCTGGCCGAC ATCCGGCCCCA CGAACACGGA CGTGGCCTTC
25551 TATTCGACGG TCACCGCCGA GCGCCTGACG GACACCACGG CCCTGGACAC
25601 GGATTACTGG GTCACCAACC TCCGTCAGCC CGTCCGGTTC GCCGACACCA
25651 TCGAAGCCCT TCTCGCGGAC GGCTACCGCC TGTTTCATCGA GGCCAGCGCC
25701 CACCCCGTGC TGGGCCTGGG CATGGAGGAG ACCATCGAGC AGGCGGACAT
25751 GCCCGCCACC GTCGTCCCCA CCCTCCGCCG CGACCACGGC GACACCACCC
25801 AGCTCACCCG CGCCGCCGCC CACGCCTTCA CCGCCGGCGC CGATGTCGAC
25851 TGGCGGCGCT GGTTCCTGGC CGACCCCGCC CCCCACACGA TCGATCTCCC
25901 CACCTACGCC TTCCAGCGCC GCCGCTACTG GCTGGCCGAC ACAGTGAAGC
25951 GGGACAGCGG ATGGGACCCG GCCGGGTCGG GGCATGCCCA GTTGCCGACC
26001 GCGGTCGCCC TCGCCGACGG GGGAGTGGTG CTGAACGGCC GGGTGTCCGC
26051 CGAGCGCGGT GGCTGGCTGG GCGGGCATGT GGTGGCGGGG ACGGTTCTGG
26101 TGCCGGGTGC GGCCTTGGTG GAGTGGGTGT TGCGGGCCGG TGATGAGGCG
26151 GGTTGCCCTT CGCTTGAGGA GTTGACGCTC CAGGCGCCGT TGGTGTGGCC
26201 CGAGTCGGGT GGGTTGCAGG TTCAGGTGGT CGTGGGTGCG GCTGATGAGC
26251 AGGGCGGCCG TCGTGACGTA CATGTGTATT CGAGGTCTGA GCAGGACGCG
26301 TCGGCGGTGT GGCAGTGCCA TGCCGTCGGT GAGCTCGGGC GCGCGTCGGT
26351 GGC GCGGCCG GTGCGGCAGG CCGGGCAGTG GCCTCCGGCG GGGGCCGAGC
26401 CGGTGGAGGT GGGCGGCTTC TACGAGGGGG TCGCGGCCGC CGGTTACGAG
26451 TACGGTCCGG CGTTCCGTGG GCTGCGCGCG ATGTGGCGGC ACGGTGATGA
26501 CCTCCTTGCG GAGGTCGAGC TGCCGGAGGA GGCCGGTTCTG CCGGCCGGTT
26551 TCGGCATCCA CCCGGCGCTG CTGGACGCCG CCCTGCACCC GCTGCTCGCA

26601 CAGCGGAGCC GGGACGGGGC CGGGGCGGGG GCCCACGGCG GGCAGGTGCT
26651 GCTGCCTTTC AGCTGGAGCG GTGTTTCCCT GTGGGCCAGC GAGGCCACCA
26701 CTGTGCGGGT GCGGCTCACC GGGCTGGGAG GAGGGGACGA CGAGACGGTG
26751 TCCCTGACGG TAACCGACCC CGCCGGTGGC CCCGTGGTGG ACGTGGCAGA
26801 GCTGCGGTTG CGGTCGACGA GCGCCCGGCA GGTGCGGGGT TCGGCAGGCC
26851 CCGGCGCGGA CGGGCTCTAC GAGCTGCGGT GGACACCGTT GCCCGAGCCG
26901 CTTCCCGTAC CGGCCCCCGC GAACGGTCGC GATGTGGCCG CCGACCTGTC
26951 CGGATGCGCG GTGCTCGGCG AACTGGTCGC GGAACCGGGC CCGGGCATCG
27001 ACCTGGAGGG CTGCCCCTGC TACCCGGGCG TCGGCGCGCT CGCCGACAAC
27051 GCCTCCCCGC CCTCCATGAT CCTCGCCCCC GTGCACAGCG ACACCACAGG
27101 CGGCGACGGA CTCGCCCTGA CGGAACGGGT GTTGCGCGTC ATCCAGGACT
27151 TCCTGGCTGC ACCGAGTCTG GAACAGAAAC AGACGCGCCT GGCCTTCGTG
27201 ACCCGGGGCG CGGCGGACAC AGGTAGCACG ACGGGAGGCT CGGCTGCCCC
27251 GGCAGAGGCA GTCGACCCGG CGGTCGCGGC CGTATGGGGC CTAGTACGCA
27301 GCGCGCAGTC GGAGAACCCC GGCCGCTTCG TACTGCTGGA CACCGACGCG
27351 CCCCTCGACC AGGCGTCCGT TGCCCCCTCTC GTGGACGCGG TCGGGTCTGC
27401 CGTGGAGGCG GACGAGCCCC AAGTCGCCCT GCGCGGGGGA CGGTTGCTCG
27451 TGCCCAGGTG GGC GCGGGGCC GGCGAGCCCG TCGAGCTGGC CGGGCCGGCC
27501 GGAGCGCGGG CGTGGCGGCT GGTGGGCGGA GACTCCGGGA CGCTGGAGGC
27551 CGTCGTGGCG GAGGCTTGCG ACGACATTGT GCTGCGCCCG TTGGCGCCGG
27601 GCCAGGTCCG CGTCGCCGTC CACACGGCCG GGGTCAATTT CCGTGACGTC
27651 CTGATCGCCC TGGGCATGTA CCCGGACCCG GACGCGCTGC CCGGCACCGA
27701 GGCGGCCGGC GTGGTGACGG AGGTCGGGCC GGGCGTCACC CGTCTGTCCG
27751 TGGGCGACCG CGTGATGGGC ATGATGGACG GCGCCTTCGG CCCGTGGGCC
27801 GTCGCCGACG CGCGCATGCT GGCCCCGGTC CCGCCCGGCT GGGGCACCCG
27851 GCAGGCGGCC GCCGCTCCCG CCGCGTTCCT GACGGCTTGG TACGGGCTGG
27901 TGGAGCTGGC CGGTCTGAAG GCGGGCGAGC GTGTGTTGAT CCATGCCGCC
27951 ACGGGTGGTG TGGGGATGGC GGCGGTGCAG ATCGCCCGGC ATGTGGGTGC

28001 CGAGGTGTTT GCCACCGCGA GTCCGGGCAA GCACGCCGTG CTGGAGGAGA
28051 TGGGCATCGA CGCCGCCCAC CGCGCCTCGT CGCGCGACCT CGCCTTCGAG
28101 GACGCCTTCC GGCAGGCCAC CGACGGCCGT GGCCTGGACG TCGTCCTCAA
28151 CAGCCTCACC GGTGAACTGC TCGACGCGTC CCTGCGATTG CTCGGCGACG
28201 GCGGGCGCTT CGTGGAGATG GGCAAGAGCG ATCCGCGCGA CCCCAGCTG
28251 GTCGCGCTGG AGCACCCCGG GGTGTCGTAC GAGGCCTTCG ACCTCGTCGC
28301 CGACGCCGGG CCCGAGCGGC TCGGGCTGAT GCTCGACAGG CTCGGCGAGC
28351 TCTTCGCCGG CGGATCACTG GTACCGCTGC CGGTCACCGC ATGGCCGCTG
28401 GGGCGGGCGC GAGAGGCGCT CCGCCACATG AGTCAGGCGA GGCACACCGG
28451 CAAGCTGGTG CTCGACGTGC CCGCGCCGCT CGACCCCGAC GGCACCGTCC
28501 TCGTCACCGG GGGTACCGGC ACCATCGGCG CCGCCGTGGC CGAACACCTG
28551 GCGCGTACCG GGGAGAGCAA GCACCTGCTC ATCGTCAGCC GCAGCGGGCC
28601 GGCCGCCCAC GCGCGCGAGG AACTTGTCTC TCGTATAGCC GAGTTCGGGG
28651 CCGAAGCCAC CTTCGTGCTG GCGACGTGA GTGAGCCCGA CGCGGTGCGC
28701 GCCCTGATCG AAGGGATCGA TCCGGCCCAT CCGCTGACCG GTGTCTGTGA
28751 TGCCGCCGGA GTACTCGACA ACGCTCTGAT CGGCTCCCAG ACCACCGAAA
28801 GCCTCACCCG CGTATGGGCG GCGAAGGCCG CCGCCGCGCA GCAACTCCAC
28851 GAGGCCACGA GGGAGTCGAG GCTGGGACTG TTCGTGATGT TCTCCTCCTT
28901 CGCCTCCACC ATGGGCACCC CAGGGCAGGC CAACTACTCC GCCGCCAACG
28951 CCTATTGCGA CGCGCTGGCC GCTCTCCGAC GCGCGGAGGG GCTCGCCGGC
29001 CTGTCCGTGG CGTGGGGGTT GTGGGAGGCC ACCAGCGGCC TGACCGGGAC
29051 GTTGTGCGCG GCCGACCGGG CCCGCATCGA CCGGTACGGC ATCAGGCCGA
29101 CCAGCGCGGC ACGCGGCTGC GCCCTGCTGG CAGCGGCACG CGCCACGGG
29151 CGCCCCGACC TGCTCGCCAT GGACCTGGAC GCCCGCGTAC CCGCCGCGTC
29201 CGACGCTCCG GTCCCCGCGG TGCTGCGCAC TCTGGCGGCC GCCGGAGCGC
29251 CCGCCACCGC CCGTCCCACC GCGGCGGCGG CCGCTGACGG GGCGACGGAC
29301 TGGTCCGGCA GGCTCGCCGG CCTCACCGAG GAGGCACGGC TCGAACTCCT
29351 CACCGAGTTG GTGTGCACCC ACGCGGCAGG GGTGCTCGGG CACGCCGACG

29401 CGGGCGCGGT CCAGGTGGAC GCGCCGTTCA AGGAACTCGG CTCGACTCG
29451 CTGACCGCCG TCGAACTGCG CAACCGGATC GCCGCCGCGA CCGGCCTGAA
29501 ACTGCCCCGC GCCCTCGTCT TCGACTACCC GCAGGCTCGC GTTCTCGCCC
29551 CCCACCTGGC CGAACGGCTC GTCCCGGAGG GCGCGGGGGC CATGGGCGGT
29601 GTGAGCGGTG CGGAGGGCGT GAGGGACGCG TACGGGGCAG GCGGTCCGGG
29651 CGGCACATG ACCGCCAGG TCTTGCTGGA GGTGGCCCGC GTCGAGCACA
29701 CCCTGTCCGC CGCCGTCCCG CACGGCCTGG ACCGGGCGGC CGTCGCGGCC
29751 CGCCTGGAGG CGCTGCTCGC CCGCTGCACG GCGACGACGG CGGCCACGGG
29801 GGCCGCGGGA GCCGCGGTGG AGGGTGACGG CGACAGCGAC GGCGACGGCG
29851 CCGTGATCA GCTGGAGACG GCCACCGCCG AGCAAGTACT GGACTTCATC
29901 GACAACGAAC TCGGGGTGTG AGCCGCGTGC CGGCCGCACA CCAGGCGATC
29951 ACGGGCGGGG AGCTGCAGCG CACATGGTGA GCGAAGAGAA ACTGGTCGAC
30001 TACCTCAAGC GTGTCTCCGC GGACCTGCAC GCCACCCGGC AGCGGCTGCG
30051 CGAGGCGGAG GAGCGCGGCC AGGAACCCGT GGCCGTGGTG GAGGCCGCCT
30101 GCCGCTACCC CGGCGGCATC CGCACCCCG AAGACCTGTG GGACCTGGTC
30151 GCCGCGGGCG GCAACGCCCT GGGCGCCTTC CCCGACAACC GCGGCTGGGA
30201 CCTGCGACGC CTCTTCACC CCGACCCCGA CCACCCCGGG ACGACCTACG
30251 CCCGCGAGGG CGGCTTCCTC CACGACGCCG ACCTGTTCGA CCCGAGTTC
30301 TTCGGCATCA GCCCCGCGA GGCCGCGGTC CTCGACCCGC AGCAGCGACT
30351 GCTCCTGGAG TGCGCCTGGG AGGCACTGGA GCGCGCGGGC ATCGACCCGC
30401 GGTCCCTCCA GGGCAGCCGT ACCGGCGTGT ACGCGGGTGC CGCCCTGCCC
30451 GGCTTCGGCA CCCCGCACAT CGACCCCGCC GCCGAGGGCC ACCTGGTCAC
30501 CGGCAGCGCC CCGAGCGTCC TCTCGGGCCG GCTCGCCTAC ACCTTCGGCC
30551 TCGAAGGGCC CGCGGTGACG ATCGACACCG CCTGCTCGTC GTCGCTCGTC
30601 GCCGTGCACC TGGCGGCCCA CGCGCTGCGG CAGCGCGAGT GCGATCTGGC
30651 GCTCGCGGGC GGTGTCACCG TCATGACCAC CCCGTACGTG TTCACCGAGT
30701 TCTCGGCCA GCGCGGCCTG GCCGCCGACG GCCGGTGCAA GCCCTTCGCG
30751 GCCGCCGCGG ACGGCACGGC CTTCTCCGAG GGCGCCGGAC TCCTCGTACT

30801 GGAACGCCTC TCCGACGCCC GCCGGGGCCGG CCACCGGGTG CTCGCCGTCA
30851 TCCGCGGCTC GGCCGTCAAC CAGGATGGCG CGAGCAACGG CCTCACCGCC
30901 CCCAACGGCC CCGCCCAGCA GCGCGTGATC CGCGCCGCCC TCGCCGGGGC
30951 GCGGCTCTCG CCCGCGGAGG TGGACGCGGT CGAGGCGCAC GGCACCGGCA
31001 CCCGGCTGGG CGACCCCATC GAGGCCGACG CGCTCCTCGC CACCTACGGT
31051 CAGGAGCGCC ACGGGGGCCG GCCGCTGTGG CTCGGCTCGG TGAAATCCAA
31101 CATCGGCCAC ACGCAGGGCG CGGCCGGTGC CGCGGGCCTG ATCAAGATGG
31151 TCCAGGCACT GCGGCACGAG ACCTGCCCCG CCACGTTGTA CGCCGACGAG
31201 CCCACCCCGC ACGCCGACTG GGAGTCGGGC GCGGTGCGCC TGCTCAGCGC
31251 GCCGGTCGCC TGGCCGCGCG GGGAGCACGG GGAGCACACC CGCAGGGCCG
31301 GCATCTCCTC CTTCCGCATC TCCGGCACGA ACGCCACCT CATCCTGGAG
31351 GAGGCGCCCG CGGCCGACGC CGAAGGAGCG GGTGGCGACG GCGATGGCGA
31401 CGGGGGAGGG GTGCGGCCGG TGGTGCGGGT CGGCGCCACG GGCCCCCGCG
31451 AAGAGCAGGG CCAAGGACAG GGCCAAGAGC AGCACCAACA GCAACGTCAG
31501 CAGCGGCAGC GGTCGTTCGAT GATGCCGACG CCGCACCTCC CGTGGCTGCT
31551 GTCCGCCCGC AGCCCCGCCG CGCTCCGCGC CCAGGCCGAC GCGCTGGCGA
31601 ACCATGTTCG CCACGCGGAC CACTCCATCG CCGACATCGG CGGCACACTG
31651 CTGCGCCGCA CCCTGTTCTGA GCACCGGGCG GTCGTCCTCG GAACCGACCG
31701 TGATGAGCGT GCCGCAGCGC TTGCCGCCCT CGCGGCAGGA CGCGCACACC
31751 CCGCGCTGAC CCGGGCCGCA GGGCCGGCGA GGAACGGCGG CACCGCCTTC
31801 CTGTTACCG GCCAGGGAAG CCAACGCCCA GGCATGGGCA GGCAGTTGTA
31851 CGACACCTTC GACGTCTTCG CCGAGTCGCT CGACGAGACC TGCGCCCGGC
31901 TCGACCCCT GCTCGAACAG CCGCTGAAGC CCGTCCTGTT CGCCCCCGCC
31951 GACACGCGC AGGCCGCGT GCTGCACGGG ACCGGCATGA CGCAGGCCG
32001 GCTGTTTCGCC CTCGAAGTCG CCCTGTACCG CCAGGTCACC TCCTTCGGGA
32051 TCGCCCCAG CCACCTGACC GGGCACTCCG TCGGCGAGAT CGCCGCCGCC
32101 CACGTCGCCG GGGTGTTCTC CCTGGCGGAC GCCTGCACGC TGGTCGCGGC
32151 CCGGGGCCGC CTCATGCAGG CCCTGCCCCG AGGTGGCGCC ATGCTCGCCG

32201 TCCAGGCGGC CGAGGACGAC GTACTGCCGC TGCTCGCCGG GCAGGAGGAA
32251 CGTCTCTCCC TCGCCGCCGT CAACGGCCCC ACCGCCGTCG TCGTGTCGG
32301 TGAGGCCGCT GCCGTCGGGG AGGTGGAGAA GCGCTGCGC GGGCGCGGAC
32351 TGAAGACCAA GCGGCTCAAC GTCAGTCACG CCTTCCACTC GCCGCTCATC
32401 GAGCCGATGC TCGACGACTT CCGCGAAGTG GCCCGCGGGC TGACCTTCCA
32451 CGCGCCGACG CTGCCCGTCG TCTCCAACCT CACCGGCCGC CTCGCCGACG
32501 CGGAGCTGAT GGCCGACGCC GAGTACTGGG TGCGGCACGT ACGCCGGCCC
32551 GTGCGGTTCC ACGACGGGCT GCGCGCTCTC AGCGAGCAAG GCGTCGTGCG
32601 CTACCTGGAG TTGGGGCCCG ACCCGGTCCT CGCCACCATG GTCCAGGACG
32651 GTCTCCCGGC CCCGGCGGAG GGAGAGGAGC CCGAGCCGGT CGTCGCCGCG
32701 GCGCTGCGCT CCAAGCACGA CGAGGGACGC ACCCTGCTGG GTGCCGTCGC
32751 CGCGCTCCAC ACCGACGGAC AGCCGGCCGA CCTCACC GCCC CTCTTCCCCG
32801 CCGACGCCGG GCAAGTGCCG CTCCCCACCT ACCGGTTCCA GCGGCGACGG
32851 TACTGGCGCG TCGCGCCCGA CGCCGCCGCG CCGGCCCGCG CCGCCGGCCT
32901 CCAGGAGACC GGCCACCCGC TGCTGCCCCG CGTCATCCGG CAGGCCGACG
32951 GCGGCATCCT GCTCGCGGGA CGCCTGTCCC TGCCTACGCA TCCATGGCTC
33001 GCCGACCACA CCATCGCGGG CGGCGTCCCG CTGCCCCGCA CCGCCTTCGT
33051 CGAACTCGCC CTGCTCGCAG GGCGGCACGC CGCCTGCGAC ACGATCGACG
33101 ATCTGACGCT GGAGACGCCG CTGCTGCTCG ACGACACCGG TACCGGTGTC
33151 GGGGCGGCTG TGGGCGCGGG CGCCGATGCC CTCGTCGATG CCATAGAAGT
33201 GCAGCTTGCC CTCGGCGCTC CCGACGGTTC CGGCCGCCGT GCTCTCACCG
33251 TCCACTCCCG TCCTGCCGAC GATGCGGCTG ACGACGGCGA CGCGGCCGAC
33301 GCGGCCGATG CGGCAGGCCG GGGAGGCCCG GGCGGCTCGG GTGACCTGGG
33351 CGATCCTGGC GATCCGGGCG ATCTGGGCGA CGGCGGGGGC TCCCGCGGCT
33401 GGCGCCGTCA CGCCACCGGC ATCTCAGCG CCGGCCCGGC CGCCGAACCG
33451 GCCGCCCCCG ACGCCGCTCC CTGGCCGCCG GCCGACGCCA CCGCCCTCGA
33501 CGTCGACGCG CTGTACGCCC GGCTCGACGC GCAGGGCTAC AGCTACGGGC
33551 CCGCCTTCCG GGCCGTCCAC GCCGCTGGC GGCACGGCGA CGACCTCTAC

33601 GCCGATGTCC GCCTCGCCGA CGAACAGCGC GCTGAAGCCG ACGCGTTCGC
33651 CCTCCACCCG GCCCTGCTCG ACGCCGCCCT GCATGCCGTC GACGAGCTGT
33701 ACCGCGGCAG TGAGGGGCGG GGGCAGGAGC AGGGGCAGGG TGGTCAGGAG
33751 CCGGAGCAGG GCCGTGGCGA CGCGGACGCC CCGGTACGGC TGCCGTTCTC
33801 CTTCAGCGAC ATACGCCACC ACGCCACCGG GGCCACACGG CTGTGGGTCC
33851 GCCTCAGCCC CCAGGGCGAC GATCGGCTGC GGCTGTCCCT GACCGACGGC
33901 GAGGGCGGGC AGGTCGCGAC AGTCGACGCC CTCCAAGTGC GGTTGATCCC
33951 CGCCGACCGG TGGCGCGCGG CCCGCCCCAC CACAGCCGCC CCCCTGTACC
34001 ACCTGGACTG GCACGAGCTG CCGTTGCCCG AGCCGGCCGA GACGGACCCG
34051 GCCGCCCCTT CCTGGGCTGT GCTCGGAGCG CACGACGCGG GCCTCGCTCC
34101 CGCCGCGCAC TACCCGGACC TGGCGGCCCT GAAAGCCGCC GTCGAGGCCG
34151 GCGAGCCCGT GCCGGACATC GTCTTCGCAC CGTTCCCCGC GCAGGGGACG
34201 GAGACCGATG TCCCGGCTCA GGTACGAGCC CACGCCCGGC ACGCCCTGGA
34251 GCTGCTGCGC GACTGGCTCA CCACGGAAGC TTTCGCCGCC GCCCGCCTCG
34301 TCGTCCTCAC GACCGGTGCG GTCACCGCCC GCCCAGAGGA CGGGCCCGCC
34351 GACCTGGCCA CCGCACCTGT ATGGGGCCTG GTCCGAGCCG CCCAGGCCGA
34401 ACAACCCGAC CATGTCGTCC TGGTGGACAT CGACAAGGAC ATCGATAAGG
34451 ACACCGACGA GGAGACCGAC CAGGCCACCG ACGCGGGCAC CGCATCGCGC
34501 CACGCTCTGC CCGCCGCCTT GGCCGCGGCG GCCGCCAAG CCGAGACACA
34551 GCTCGCCCTG CGCGCGGGCA CCGTGCTCGT GCCGCGCCTC GCCGTCGTCC
34601 CGCCCCGGAC CGACACCCCA GCGCTGCACG CCACCGCCCC GGAGAGCACC
34651 ACGGACACTG TGGACTCCAC GGGCATCGCG GGCCTGCGG AATCCGGCGG
34701 CACCGTCCTG ATCACCGGCG GAACCGGCGG CCTCGGGCAG GCCGTCGCCC
34751 GTCACCTCGC CGCCGCGCAT GGCGCCCGCC ACCTGCTCCT CGTCAGCCGC
34801 AGGGGCGACG CCGCCGAGGG CGTCGCCGAG TTGCGCGCCG ACCTCGCGGA
34851 CGACGGCGTC GACGTACGCG TCGCCGCTG CGACATCACC GACCGCGACG
34901 CGCTGGCCGG GCTCCTCGCG GACATCCCCG CCGCGCAGCC GCTCACCAGG
34951 GTCGTGCACA CCGCGGGCGT CATCGACGAC AGCCTCATCA CGGCGATGAC

35001 CCCCAGGCGG CTCGACGCCG TCCTCGCACC CAAGGCCGAC GCGGCCTGGC
35051 ACCTGCACGA ACTCACC CGC GACAAGGACC TGTCGGCCTT CGTCCTCTTC
35101 TCCTCGGGCG CCTCCGTCCT CGGCAACGGC GGCCAGGCCA ACTACGCGGC
35151 CGCCAACACC TTCCTCAACA CCCTCGCCGA ACACCGCCGC GCGGCCGGCC
35201 TCGCCGCCAC CTCCGTGGCC TGGGGCCTGT GGGAGTCCGC GTCCGGCGGC
35251 ATGGCCGCCG GGCTCGGCGA CGCCGACCGC GCGCGCATCC ACCGCACCGG
35301 CGTGACGGGC CTGACCGACG AGCAGGCCCT GGCCCTCTTC GACGCGGCCC
35351 TGACCGCCGA GCACCCACG GTCCTCGCCA CCCGCTTCGA CCGCGCCGTG
35401 CTGCGCGGCC AGGCCGCCGC CCGCACCTTG CAGCCCGCCC TGCGCGGCCT
35451 GGTACGCACT CCGCGCCCCA CCGCGTCCGC CGGGGCCATC GGGTCCACCG
35501 CAGCCACCGG GTCCGCCACG GACGAGAACG CGCCCTCCTC GTGGGCCGCC
35551 CGGCTCGCCC GGCTGTCCGC CGCCGACCGC GACCGCGCCC TCAACGAACT
35601 CATTTCGCGAG CAGATCGCGA CCGTCCTGGC ACACCCCTCA CCCGACACCA
35651 TCGAACTGGG CCGCGCCTTC CAGGAGTTGG GCTTCGACTC GCTCACCGCC
35701 CTGGA ACTCC GCAACCGCCT CTCCACGGCC ACCGGCATCC GGCTGCCCCG
35751 CACCCTCGTC TTCGACCACC CGAGCCCCAC CGCCCTCGTA CGCCATCTCC
35801 ACAGCCATCT CCCCAGCAGG GCCCAGCACA CGTCCCCGAC CGCCCCCGGC
35851 GCCTCTGCGG AGGGCACCGC CGCCACGGCC ACCGGCATCG ACGACGACCC
35901 GATCGCCATC GTCGGCATGG CGTGCCGCTA CCCGGGCGGC GTGACCTCGC
35951 CCGAGCAGCT GTGGCAGCTC GTGGCCACCG GCACCGACGC CATCGGCCCG
36001 TTCCCCGAGG ACCGCGGCTG GGACACGGCC GGACTGTTCG ATCCCGACCC
36051 CGACCAGGTC GGCCACAGCT ACACCCGCGA AGGCGGCTTC CTCTACGACG
36101 CCGCCCGCTT CGACGCGGGC TTCTTCGGCA TCAGCCCGCG CGAGGCCGCC
36151 GCCACCGACC CGCAGCAGCG CCTGCTCCTG GAAACCGCCT GGCAGGCGTT
36201 CGAACACGCG GGCATCGACC CCGCCGCCCT GCGCGGCACC CCGTGCGGCG
36251 TCATCACCGG AATCATGTAC GACGACTACG GATCCCGCTT CCTCGCGCGC
36301 AAACCGGACG GCTTCGAGGG CCGCATCATG ACCGGCAGCA CGCCGAGCGT
36351 GGCCTCCGGC CGGGTCGCGT ACACCTTCGG CCTGGAGGGC CCCGCCATCA

36401 CGGTGGACAC CGCGTGCTCC TCCTCGCTGG TCGCGATGCA CCTGGCGGCG
36451 CAGGCGCTGC GGCAGGGCGA GTGCGAACTG GCCCTGGCCG GGGGTGTGAC
36501 CGTGATGGCC ACCCCGAACA CCTTCGTGGA GTTCTCCCGC CAGCGCGGCC
36551 TGGCCCCCGA CGGCCGCTGC AAGCCGTTCG CCGCCGCGGC GGACGGCACC
36601 GGCTGGGGCG AGGGCGCCGG ACTCGTCGTC CTGGAGCGCC TCTCCGACGC
36651 GCGCCGCAAG GGACACCGCG TCCTCGCCCT GCTGCGCGGT TCGGCCGTGA
36701 ACCAGGACGG CGCGAGCAAC GGCATGACCG CCCCGAACGG TCCCTCGCAG
36751 GAACGGGTCA TCCGCACCGC CCTGGCCGGC GCGGGCCGTG GTCCCAGGA
36801 CATCGACGTG GTGGAGGCGC ACGGCACCG CACCACGCTC GGCGACCCGA
36851 TCGAGGCGCA GGCCCTGCTC GCCACGTACG GGCAGGGGCG CCCGGAGGAC
36901 CGCCCGCTCT GGCTCGGCTC GGTGAAGTCG AACATCGGCC ACACGCAGGC
36951 CGCCGCCGGT GTCGCGGGCG TCATCAAGAT GGTCATGGCA CTGCGCCACG
37001 AGCAACTGCC CACGACCCTG CACGCCGACG AGCCGACCCC CCACGTGCAA
37051 TGGGACGGCG GCGGCGTACG TCTCCTGACC GAACCGGTCC CGTGGTTCGG
37101 CGGCGAGCGC ACGCGGCGCG CCGGGGTGTC GTCCTTCGGG ATCTCCGGGA
37151 CGAACGCGCA CCTGATCCTG GAGGAGCCGC CGGAGGAGGA CCTGCCCCGAG
37201 CCCGTGGCGG CGGAGCCGGG TGGGGTGGTG CCGTGGGTGG TGTCCGGGCG
37251 GACGCCGGAC GCGTTGCGTG AACAGGCGCG GCGGCTCGGC GAGTTTGTCTG
37301 TCGGTGCCGG GGATGTGTCTG GCAGCCGAGG TGGGATGGTC ACTGGCCACG
37351 ACGCGGTCTG TGTTCGAGCA CCGGGCCGTG GTGGCGGGCC GGGACCGGGA
37401 CGATCTGGTT GCCGGGATGC AGGCGCTGGC GGCAGGGGAG ACGCCGACAG
37451 ATGTCGTGTC CCGTGCGGCG GCTTCCTCCG GTGCGGGGCC GGTGTTGGTG
37501 TTCCCGGGGC AGGGGTCGCA GTGGGTGGGC ATGGGTGCCC AGCTCCTTGA
37551 CGAGTCCCCC GTCTTCGCGG CGCGGATCGC GGAGTGTGAG CAGGCGCTGT
37601 CGGCGTACGT GGA CTGGTCTG CTGAGTGATG TCCTGCGCGG GGACGGGAGT
37651 GAGCTGTCCC GGGTCGAGGT CGTGCAGCCC GTGTTGTGGG CGGTAATGGT
37701 CTCGCTGGCT GCCGTCTGGG CGGATTACGG GGTCACCTCC GCCGCTGTGG
37751 TGGGGCATTG GCAGGGTGAG ATGGCTGCCG CGTGTGTGGC GGGGGCGCTG

37801 TCGCTGGAGG ATGCGGCGCG GATTGTGGCG GTACGCAGTG ACGCGCTTCG
37851 TCAGCTGCAA GGGCACGGCG ACATGGCCTC ACTCGGCACT GGTGCCGAGC
37901 AGGCCGCTGA GCTGATCGGT GATCGGCCGG GAGTGGTCGT CGCGGCAGTC
37951 AACGGGCCGT CGTCTACCGT GATTTCGGGG CCGCCGGAGC ATGTGGCCCG
38001 TGTGGTCGCG GAGGCGGAGG CACGTGGTCT GCGCGCCCGT GTGATCGACG
38051 TCGGGTATGC CTCGCACGGC CCCCAGATCG ACCAGCTCCA CGACCTCCTC
38101 ACCGAGGGCC TGGCTGACAT CCGGCCCGCG AACACGGACG TGGCCTTCTA
38151 TTCGACGGTC ACCGCCGAGC GCCTGACGGA CACCACAGCC CTGGATACGG
38201 ATTACTGGGT GACCAACCTC CGCCAGCCGG TCCGGTTCGC CGACACCATC
38251 GAAGCGCTTC TCGCGGACGG CTATCGCCTG TTCATCGAGG CCAGCGCGCA
38301 CCCGGTGTTC GGCCTGGGCA TGGAGGAGAC CATCGAGCAG GCGGACATCC
38351 CTGCCACGGT CGTCCCCACC CTGCGCCGCG ACCACGGCGA CACCACCCAG
38401 CTCACCCGCG CCGCCGCCCA CGCCTTCACC GCCGGCGCCG ATGTCGACTG
38451 GCGACGCTGG TTCCCGGCCG ACCCCACCCC CCGTACCGTC GACCTCCCCA
38501 CCTACGCCTT CCAGCACCAG CACTACTGGC TGGAGGAGCC CAGTGGGCTC
38551 ACCGGAGACG CCGCCGACCT CGGCATGGTG GCCGCCGGGC ATCCGCTGCT
38601 CGGTGCCTGT GTGGAACCTG CGGAGAGCGA CTCGTACTTG TTCACCGGGC
38651 GGCTCTCGCG CAGGGCTCCG TCCTGGCTGG CCGAACACGT GGTGGCGGGG
38701 ACGGTTCTGG TGCCGGGTGC GGCCTTGGTG GAGTGGGTGC TGCGGGCCGG
38751 CGATGAGGCG GGATGCCCGA CGATTGAGGA ACTGACGCTC CAGGCGCCGT
38801 TGGTGCTGCC CGAGTCGGGC GGGTTGCAGG TTCAGGTGGT CGTGGGTGCG
38851 ACCGATGAGC AGAGCGGCCG TCGTGACGTA CACGTGTATT CGAGGTCTGA
38901 GCAGGACGCG TCGGCGGTGT GGGTGTGCCA TGCCGTCGGT GTGGTGAGCT
38951 CCGAAATGCC AGAAGCGGCA GCCGAGTTGA GTGGGCAGTG GCCTCCTGCC
39001 GGGGCCGAAG CCGTGGATGT CGAGGACTTC TACGCGCGGG CCGCGGAGGC
39051 CGGATACGCC TACGGTCCGG CGTTCCAGGG GCTGCGGGCG CTGTGGCGGC
39101 ACGGGACGGA GCTGTTCGCC GAGGTGGTGC TGCCCGAACA GGCGGGTGGG
39151 CACGACGGTT TCGGCATCCA CCCGGCGCTG CTGGACGCCG CCCTGCATCC

39201 GCTGATGCTC CTCGACCGGC CCGCGGACGG GCAGATGTGG CTGCCGTTTCG
39251 CGTGGAGCGG GGTGTCGCTG AACGCGGACC GGGCGACCCA CGTCCGTGTC
39301 CGGCTCTCCC CGCGGGGGGA GGCGGCCGAG CGTGACCTGC GGGTCGTCAT
39351 CGCCGACGCG ACCGGCGCGC CCGTCCTGAC GGTCGACGCC CTGACCCTGC
39401 GCGCGGCCGA TCCCGGCCGG CTGGGTGCGG CGGCCCCTGG CCGTGTCGAC
39451 GGCCTCTACA CCGTCGACTG GACCCCGCTG CCCCTGCCCC AGCCCCTTCC
39501 GCTGCCGCGG ACGGATGCAG GGGGGAGTGC CGACTGGGTC ATACTCTCGG
39551 ACAACTCCAG TGCAGCTCTG GCTGATGCCG TGTCGTCCGC GACGGCGGCA
39601 GGTGGCGGAG CGCCGTGGGC ATTGCTCGCT CCCGTGGGTG GCGGCTCTGC
39651 CGATGACGGG CTGCCGGTGG TCGGCGGAC CCTCTCCCTC GTACAGGAGT
39701 TCCTGGCCGC CCCGGAGCTG ACCGAGTCCC GTCTCGTCAT CGTGACACGC
39751 GGTGCCGTGG CCACCGACGC CGATGGTGAC GTCGCGGCGT CCGCGGCAGC
39801 GGTATGGGGC CTGATCCGCA GCGCCCAGTC GGAGAACCCG GGCCGCTTCG
39851 TCCTGCTCGA CGTCGAGGAG GAGCACCTCC ACCCGGACGG CGGGGAAGT
39901 CCGTACGCCG CCCTGCGCCA CGCCGTAGAG GAGCTCGACG AGCCTCAACT
39951 TGCCCTCCGC AGCGGCAAAT TCCTCGTACC GCGCATGACG CCCGCCGCCG
40001 CCCCCGAGGA GCTCGTCCCG CCGGTCGGTA CGTCCGGCTG GCGCCTCGGC
40051 ACCTCCGGTA CGGCCACCCT GGAGAATCTG TCGGTGATCG ACGTCCCGA
40101 GGC GTTTCGCG CCGCTGGAGC CCGGGCAGGT GCGGATCTCC GTACGGGCGG
40151 CGGGCATGAA CTTCCGTGAC GTGCTGATCG CGTTGGGCAT GTATCCCGAC
40201 AAGGGCACGT TCGCGGGAAG CGAGGGCGCC GGACATGTGA CGGAGGTGGG
40251 ACCGGGCGTC ACTCATCTGT CGGTCGGTGA CCGGGTGATG GGTCTGTTTCG
40301 AGGGCGCGTT CGCTCCGCTG GCCGTCGCGG ACGCCCGGAT GGTCGTCCCG
40351 ATTCCGGAGG GCTGGAGCTT CCAGGAGGCC GCGGCGGTGC CCGTGGTGTT
40401 CCTCACGGCC TGGTACGGCC TCGTGGACCT CGGCCGCCTC CGGGCGGGCG
40451 AATCGCTGCT CATCCACGCG GGCACCGGCG GAGTGGGCAT GGCCGCCACC
40501 CAGATCGCCC GCCACCTGGG CGCCGAGGTG TTCGCCACCG CGAGCCCCGC
40551 CAAGCACGGC GTGCTCGACG GCATGGGCAT CGACGCGGCC CACCGCGCCT

40601 CCTCCCGTGA CCTCGACTTC GAGGAGACCT TGCGGGCGGC GACGGGCGGG
40651 CGCGGCATGG ACGTCGTACT CAACAGTCTG GCCGGGGAGT TCACCGACGC
40701 CTCGCTGCGG CTGCTCGCCG AGGGCGGGCG CATGGTGGAC ATGGGCAAGA
40751 CCGACAAGCG CGACCCCGAC CGGGTCGCGG CCGAGCACGC GGGCGCGTGG
40801 TACCGGGCCT TCGACCTCGT GCCGCACGCG GGGCCCGACC GGATCGGGGA
40851 AATGCTGGCG GAGCTGGGCG AGTTGTTCGC CTCCGGCGCC CTGGCGCCGC
40901 TGCCCGTCCA GACCTGGCCG CTGGGCCGGG CGCGTGAGGC GTTCCGGTTC
40951 ATGAGCCAGG CGAAGCACAC CGGCAAGCTG GTGCTGGAGA TCCCGCCCGC
41001 CCTCGATCCG GACGGCACGG TGCTCATCAC CGGCGGCACC GGGGTCCTCG
41051 CCGCCGCGGT GGCCGAGCAT CTGGTGAGGG AGTGGGGCGT ACGACACCTG
41101 CTGCTGGCCG GGAGGCGCGG TTCCGAGGCG CCCGGGAGCA GTGAACTCGC
41151 CGAGGAACTG ACCGAGTTGG GGGCCGAGGT GACCTTTGCC GCGGCCGATG
41201 TCAGTGATCC GGACGCCGTG GCGGAGCTCG TCGGCAAGAC CGATCCGGCG
41251 CACCCGCTGA CCGGTGTGAT CCACGCGGCC GGTGTGCTGG ACGACGCCGT
41301 GGTCACCGCA CAGACCCCGG AGAGCCTCGC GCGGGTGTGG GCGGCGAAGG
41351 CGACGGCCGC ACACCTGCTG CACGAGGCGA CCCGGGAGGC GCGCCTCGGT
41401 CTCTTCCTGG TGTTCCTCCTC GGCGGCGGCG ACACTCGGCA GTCCGGGACA
41451 GGCCAACTAC GCGGCGGCCA ACGCCTATTG CGACGCCCTC GTCCGGCAAC
41501 GGCGTGCCGA GGGCCTGGCC GGTCTCTCGA TCGGCTGGGG TCTGTGGCAG
41551 ACGGCGAGCG GCATGACCGG ACACCTCGGC GAGACGGACC TGGCACGCAT
41601 GAAGCGCACC GGGTTCACCC CGCTGACCAC CGAAGGTGGC TTGGCCCTCC
41651 TCGACGCCGC CCGCGCCAC GGCCGCCCGC ACGTGGTCGC GGTGGACCTC
41701 GACGCGCGCG CCGTCGCCGC GCAGCCCGCC CCGTCCCGGC CCGCGCTCCT
41751 GCGCGCCCTG GCCGCGGGTG CGACCCCGGG GGCCCGCACC GCCCGGCGCA
41801 CCGCGGCCGC GGGCAGCGTC GCCCGGCGG GCGGTCTCGC CGACCGGCTC
41851 GCCGGCCTGC CGCATCCCGA ACGGCGCCGG CTGCTGCTCG ACCTCGTACG
41901 TGGCAACGTC GCCGGCGTCC TCGGGCACAG CGACCACGAC GCCGTCCGCC
41951 CGGACACGTC GTTCAAGGAG CTCGGCTTCG ACTCCCTGAC CGCCGTGGAA

42001 CTGCGCAACC GGCTGGCCGC CGCCACCGGC CTGAAGCTGC CCGCGGCGCT
42051 CGTCTTCGAC TACCCCGAGT CGGCCACCCT CGTCGACCAC CTCCTGGAGC
42101 GTCTGTGCGC CGACGGCGCG CCGCCGCCCG TCAAGGACGC CGCGGACCCC
42151 GTTCTCAACG ACCTCGGCAG GATCGAGTCC TCCCTGGACG CGCTCGCCCT
42201 CGACGCGGAC GCGCGCAGCC GGGTCACCAG GCGTCTGAAC ACCCTGCTGT
42251 CGAAGCTGAA CGGAGCCGCC ACCGCCGGCT CCCC GGCGGA CGTCACGGAC
42301 CTGGACGCGC TGGACGCGCT GGACGACGTG TCCGACGACG AGATGTTCGA
42351 GTTCATCGAC CGAGAGCTGT GACCCCCCTG CCGCCCCGT CCCCCTTCCC
42401 CGCCCCCAGC TTCCCCGTGC CCTTCGCTGA TGGAGAAGTG ACGTTCGATG
42451 TCGAGTGCTG AAGAGTCGAG TCCTGATGTG TCCGGCACGG GTGTGTCCGG
42501 TACGGGAGAG TCCGCTACGG GTACGTCGAG TACGGAAGCC AAGCTTCGGC
42551 AGTATCTGAA GCGGGTCACG GTGGACCTCG GCCAGGCCCG CCGGCGGCTG
42601 CGCGAGGTGG AGGAGCGGGC CCAGGAGCCG ATCGCCATCG TCTCCATGGC
42651 GTGCCGCTTC CCCGGCGACA CCCGCACGCC CGAGGCCCTG TGGGACCTGG
42701 TCGCCGAGGG CGGCGACGCC ATCGACGACT TCCCCACCAA TCGCGGCTGG
42751 GACCTGGAGA GCCTCTACCA CCCC GACCCC GACCACCCG GCACCAGCTA
42801 CGTCCGACGC GCGGGGTTCC TGTACGACGC CCGCGCCTTC GACGCGTCGT
42851 TCTTCGGGAT CAGCCCGCGC GAAGCCCTGG CCATGGACCC GCAGCAGCGG
42901 GTGCTCATGG AGACGGCCTG GCAGCTCCTG GAGCGGGCCG GCATCGACCC
42951 GGCCTCGCTG AAGCTGAGCG CCACCGGCGT CTACATCGGC GCGGGCGTGC
43001 TCGGGTTCGG CGGCGCGCAG CCCGACAAGA CGGTAGAGGG CCACCTCCTG
43051 ACCGGCAGCG CGCTGAGTGT CCTGTCCGGC CGCATCTCCT TCACGCTCGG
43101 CCTCGAGGGC CCGTCGGTCA GTGTCGACAC GGCCTGCTCC TCCTCGCTGG
43151 TCTCCATGCA CCTGGCGGCC CAGGCGCTGC GGCAGGGGGA GTGCGATCTC
43201 GCGCTGGCCG GCGGTGTCAC CGTGATGTCG ACGCCCGGCG CGTTCACCGA
43251 GTTCTCCCGC CAGGGCGCGC TGTCTCCGGA CGGCCGCTCG AAGGCTTTTCG
43301 CGGCCTCGGC CGACGGCACC GGTTTCTCGG AGGGCGCGGG ACTGCTCCTC
43351 CTGGAGCGGC TCTCCGACGC GCGCCGCAAC GGCCACAAGG TGCTCGCGGT

43401 GATCCGCGGC TCGGCCGTCA ACCAGGACGG CGCGAGCAAC GGTCTCACCG
43451 CCCCCAACGG CCCCTCCCAG GAACGCGTGA TCCGCGCCGC CCTCGCCAAC
43501 GCGGGCCTGG GCGCCGCCGA GGTGACGCG GTCGAGGCAC ACGGCACCGG
43551 CACGAAGCTC GGCACCCCCA TCGAGGCCGG TCGCTGCTC GCCACCTACG
43601 GCCGCGACAG GGACGAGGAC CGGCCGCTGT GGCTGGGCTC GGTCAAGTCG
43651 AACATCGGTC ACCCGCAGGG CGCAGCAGGC GTCGCGGGCG TCATCAAGAT
43701 GGTGATGGCG CTGCAGCGCG AACTGCTCCC CGCCACCCTG TACGTCGACG
43751 AGCCACCCCC GCACGTCGAC TGGTCCTCGG GCTCCGTCAG GCTCCTCACC
43801 GAACCGGTCC CGTGGACCCG CGGCGAGCGC CCGCGCCGCG CGGGCGTGTC
43851 CGCCTTCGGC ATGTCCGGGA CGAACGCCCA CGTGATCCTG GAGGAGGCAC
43901 CGCCCGAGGA GGCAGCGGCC GCGGAGACAC CGGCGGAAGG GACAGGCGCA
43951 GTCGTCCCGT GGGTCGTCTC CGGCCGGGGC GAGGAAGCGC TCGGGGCCCA
44001 GGCCGCACAG CTCGCCGAGC ACGTGC CGA CGACGACCAG CGGCCGGCGT
44051 CACCGCTGGA GGTGGGGTGG TCGCTCGCCA CGACACGGTC GGTGTTTCGAG
44101 AACC GGGCCG TCGTCGTCGG GGACGACCGC GACGCGCTCC TCGACGGCCT
44151 CCGGTCGCTG GCGGCAGGTG AGGCGTCGCC GGACGTGGTG TCCGGGGCGG
44201 TCGGCCCCAC GGGGCCCGGG CCGGTCATGG TGTCCCCGG CCAGGGCGGC
44251 CAGTGGGTGG GCATGGGGGC CCGGCTCCTC GACGAGTCCC CGGTGTTCGC
44301 GGCCCGGATC GCCGAGTGC AGCAGGCCCT GTCGGCGTAC GTGGACTGGT
44351 CCCTGACCGA CGTGCTGCGC GGGGACGGGT CGGAGCTGGC CCGGATCGAC
44401 GTCGTCCAGC CCGTGCTGTG GGCCGTCATG GTCGCGCTCG CCGCCGTCTG
44451 GGCGGACCAG GGAATCGAAC CCGCCGCCGT CGTCGGCCAC TCGCAGGGCG
44501 AGATAGCCGC GGC GTGCGTC GTGGGCGCCA TCTCCCTGGA CGAGGCGGCC
44551 CGCATCGTCG CCGTACGCAG TGTGCTGCTG CGGCAGCTGT CCGGACGCGG
44601 CGGCATGGCG TCCCTGGGGA TGGGCCAGGA GCAGGCCGCC GACCTGATCG
44651 ACGGACACCC GGGTGTGGTC GTCGCGGCCG TCAACGGGCC GTCGTCCACC
44701 GTCATCTCGG GCGCGCCGA GGGCATCGCC GCCGTCGTCG CCGACGCCCA
44751 GGAGCGGGGC CTTCGCGCCA GGGCCGTCGC CTCCGACGTC GCGGGCCACG

44801 GCCCGCAGCT GGACGCGATC CTGGACCAGC TCACGGAGGG CCTGGCCGGC
44851 ATCCGGCCCC CCGCGACCGA CGTCGCGTTC TACTCCACCG TCACCGCCGG
44901 GCACCTCACC GACACCACCG AACTCGACAC CGCGTACTGG GTGCGGAACG
44951 TGCGCCGGAC GGTGCGTTTC GCCGACACGA TCGACGCGCT GCTCGCGGAC
45001 GGGTACCGCC TGTTTCATCGA GGTGAGCCCC CACCCCGTCC TCAACCTCGC
45051 GCTGGAAGGC CTCATCGAAC GGGCGGCCGT GCCCGCCACG GTCGTGCCCA
45101 CCCTGCGCCG CGACCACGGC GACACCACCC AGCTCGCCCG CGCCGCGGCC
45151 CACGCCTTCG CCGCCGGCGC GGACGTCGAC TGGCGGCGCT GGTTCCTGGC
45201 CGACCCCGCC CCCCCTACCG TCGACCTGCC CACCTACGCC TTCCAGCGCC
45251 AGGACTTCTG GCCGGCCCCC GCCGGCGGGC GGTCCGGCGA CCCTGCCGGG
45301 CTCGGCCTCG CCGCCTCCGG ACACCCGCTC CTGGGCGCCT CCGTGGGCCT
45351 CGCGAGCGGG GACGTACACC TGCTGAGCGG GCGGGTGTCC CGGCAGTCCG
45401 CCGCGTGGCT GGACGACCAC GTCGTGGCGG GCCAGGCCCT GGTGCCCCGGC
45451 GCGGCGCAGG TGGAGTGGGT GCTGCGGGCC GCGACGACG CGGGCTGCTC
45501 CGCCCTGGAG GAGCTGACGC TCCAGACGCC GCTCGTGCTG CCCGACACCG
45551 GCGGCCTGCG GATCCAGGTC GTCGTGCAAG CGGCCGACGC ACACGGCCGG
45601 CGCGACGTCC GGCTGTTCTC CCGCCCCGAT GACGACGACG CTTTCGCGTC
45651 GACGCACCCC TGGACCTGCC ACGCCACGGG CGTGCTCGCC CCCGCCCCGA
45701 CGGACGGCAC CAACGGAACG CGGGACGCCG CCGACACCCT GGACGGCGCA
45751 TGGCCCCCGG CCGACGCCGA ACCCGTCCCC GCCGACGACC TCTACGCGCA
45801 GGCCGACCGC ACCGGATACG GCTACGGCCC CGCCTTCCGG GGCCTACGGG
45851 CGCTGTGGCG CCACGGCAAG GACGTCCTGG CCGAGGTGAC GCTGCCCAAG
45901 GAGGCCGGCG ACCCGGACGG CTTCGGTATC CACCCGGCCC TCCTCGACGC
45951 CGTCCTGCAA CCCGCCGCAC TGCTGCTGCC CCCGACCGAC GCCGAACAGG
46001 TCTGGCTGCC GTTCGCCTGG AACGACGTGG CGCTGCACGC CGTACGGGCC
46051 ACCACGGTCC GGGTGCGCCT CACCCCGCTC GGCGAGCGGA TCGACCAGGG
46101 GCTGCGCATC ACCGTGGCCG ACGCCGTGGG CGCGCCCGTG CTCACCGTCC
46151 GCGACCTGCG CTCGCGCCCG ACCGACACAG GCCGCCTCGC CGCGGCCGCG

46201 ACCCGCGACC GGCACGGGCT GTTCGACCTG GAGTGGATCG CGCCGGAGAA
46251 CGCGGCGGAG AACGCGGCGG GTCCGGCCCCG GGACGCGTCC GAAGGGTGGG
46301 TGACACTCGG CGAGGACGCC GCGAGCCTCG CGGACCTGCT GCGTCCGTC
46351 GAGGCGGGCG CTCCGGCGCC GCAGCTCGTG GCCGCCCCCG TCGAACCCGA
46401 CCGGACCGAC GACGGCCTGG CACTCGCCAC CCACGTCTC GACCTCGTAC
46451 AGACCTGGCT CGCCTCGCCC CTGCACGACT CCCGCTGGT CCTGGTGACG
46501 CGAGGGGCAG TGACGGATGC GGATGTGGAT GTGGCTGCCG CGGCCGTTT
46551 GGGTCTGGTA CGCAGCGCCC AGTCGGAGCA CCCCGGCCGC TTCACGCTGA
46601 TCGACCTCGG CCCCACGAC ACGCTTGCCG CAGCCATGCA GCGGGCGCAC
46651 CTGGAAGAGC CGCAACTGGC GGTGCACGGC GCGAGATAC GAGTGCCGCG
46701 ACTGGTCCGC GCCACGACCG ACCCGACCGC CCCGAACGGG ACACCGGAGG
46751 CCGACCGGAC GCGGGACCCG TCCGAAGGAC TCCACCGGAA CGGTACGGTT
46801 CTCATCACCG GCGGCACCGG CGTACTCGGC CGACTGGTGG CCGAACACCT
46851 GGTCACGGAG TGGGGCGTAC GCCACCTGCT GCTCGCGAGC CGACGCGGCG
46901 ACCAGGCGCC GGGTAGCGCC GAACTCCGCG CCCGCCTGAG CGAATTGGGA
46951 GCATCGGTCT AGATCGCCCC GGCCGATGTC GCGACGCGG AAGCGGTCTG
47001 CGCACTGATC GCGTCGGTCG ACCCGGCGCA CCCGCTCACC GGTGTGATCC
47051 ACGCGGCCCG TGTCCTGGAC GACGCCGTGA TCACCGCCCA GACCCCCGAG
47101 AGCCTCGCGC GGGTGTGGGC GACGAAGGCG ACGGCGGCCC GCCATCTGCA
47151 CGAGGCGACA CGGGAGACAC CCCTCGACTT CTTGCTGGTG TTCTCCTCGG
47201 CGGCCGCCCTC GCTCGGCAGC CCCGGCCAGG CCAACTACGC GCGGGCCAAC
47251 GCCTATTGCG ACGCCCTCGT CCAGCACCGC CGCGCCCAAG GGCTCGCGGG
47301 CCTCTCGATC GCCTGGGGCC TGTGGCAGGC GACCAGCGGC ATGACCGGGC
47351 AGCTGAGCGA GACCGACCTG GCGCGCATGA AGCGCACCGG GTTCGCCGCG
47401 CTGACCGACG AGGGCGGCCT GGCCCTGCTC GACGCCGCCC GTGCCACGA
47451 CCGGGCCTAC GTGGTCGCGG CCGACCTCGA CCCGCGCGCC GTGACCGATG
47501 GCCTGTCCCC GCTCCTGCGC GCCCTCACGG CGCCCGCCAC GCGGCGGCGC
47551 GTGGCCTCCG AAGGCCTCGC CGACGGGGCG CTCGCGACCC GCCTGGCCGG

47601 CCTCGACGCG GACGGCCGCC TAAGGCTCCT CACCGATGTC GTACGCGAGT
47651 ACGTCGCGGC CGTCCTCGGC CATGGTTCCG CCGCCCGGGT GGGCGTCGAC
47701 ATCGCCTTCA AGGACCTGGG TTTCGACTCG CTGACCGCGG TGGAGCTGCG
47751 CAACCGGCTG TCGGCCGCCT GTGACGTGCG GCTGCCCCGCC AACTGATCT
47801 TCGACCACCC CACCCCGCAG GCTCTCGCCA CCCACCTGGT GGACCGCTTG
47851 GCGGGCAGCA CCTCCGCGAC CACGACGGTG AATGCGACGG CGCCGGCAGC
47901 CGCCACGTC GCCGCAGGG CCGACGTCGA CGCAGACACC GACGACCCGG
47951 TCGCCATCGT CGCCATGACG TGCCGGTTCC CGGGCGGCGT CGCGTCCCCG
48001 GACGACCTGT GGGACCTGCT CGACGCACGC AAGGACGCGA TGGGCGCCTT
48051 CCCCACCGAC CGCGGCTGGG ACCTGGAACG CCTCTTCCAC CCCGACCCGG
48101 ACCACCCCGG CACCAGCTAC ACCGACCAGG GCGGATTTCT TCCCGACGCG
48151 GGTGATTTCG ATGCGGCGTT CTTGCGGATC AATCCGCGGG AGGCGCTGGC
48201 GATGGATCCG CAGCAGCGGT TGTTGCTGGA GCGGTCGTGG GAGGTGTTGG
48251 AGCGTGCGGG TATCGATCCG ACGACGCTCA AGGGCACCCC GACCGGCACC
48301 TACGTGGGCC TCATGTACCA CGACTACGCC AAGTCCTTCC CCACGGCCGA
48351 CGCCCAGTTG GAGGGCTACT CCTACTTGGC GAGCACCGGC AGCATGGTCT
48401 CCGGCCGCGT CGCCTACACC CTGGGCCTTG AAGGTCCGGC GGTGACGGTC
48451 GACACCGCGT GCTCCTCCTC CCTGGTCTCC ATCCACCTGG CGACGCAGGC
48501 ACTCCGGCAC GGCGAGTGCG ACCTCGCCCT GGCAGGCGGT GTGACCGTCA
48551 TGGCCGACCC GGACATGTTT GCGGGCTTCT CGCGCCAGCG CGGCCTCTCA
48601 CCTGACGGCC GCTGCAAGGC CTACGCCGCC GCGGCCGACG GAGTCGGATT
48651 CTCCGAGGGA GTGGGCGTAT TGCTCCTTGA GCGGTTGTCG GATGCGCGGC
48701 GTCATGGGCG TCGGGTGTTG GGTGTGGTGC GGGGTTCCGC GGTGAATCAG
48751 GACGGTGCGA GTAATGGGTT GACGGCGCCG AATGGTCCGT CGCAGGAGCG
48801 GGTGATTTCG CAGGCGTTGG CCAGTGGTGG GTTGTCGTCG GTGGATGTTG
48851 ATGTGGTGGA GGGGCATGGG ACGGGGACCA CGTTGGGTGA TCCGATCGAG
48901 GCGCAGGCTC TGCTGGCCAC ATATGGGCAG GGGCGTCCGG AGGACCGTCC
48951 GTTGTGGTTG GGGTCGGTGA AGTCGAACAT TGGTCATACG CAGGCGGCTG

49001 CGGGTGTTGC GGGTGTGATC AAGATGGTGA TGGCGATGCG GCATGGTGTG
49051 GTGCCGGCGA GTTTGCATGT GGATGTGCCG TCGCCGCATG TGGAGTGGGA
49101 TTCGGGTGCG GTGCGGTTGG CGGTTGAGTC GGTGCCATGG CCGCAGGTGG
49151 AGGGTCGTCC GCGTCGGGCG GGTGTGTCGT CGTTCGGCGC TTCGGGGACG
49201 AATGCGCACG TGATCGTGGA GTCTGTTCCC GATGGGCTGG AGGAGGACTC
49251 GGTATCGGTC GGCGGTGAGG CTCTGAGAC GGAGACTGAC GGGCGCTTGG
49301 TGCCGTGGGT GGTGTCGGCC CGCAGCCCGC AGGCCCTGCG CGACCAGGCA
49351 CTACGCCTGC GTGACTTTGC CAGTGACGCG TCGTTCCGCG CGCCGCTCGC
49401 CGACGTGGGC TGGTCGCTGC TGAAGACGCG TCGCTGCAT GAGCATCGCG
49451 CCGTTGTGGT GGGCGCGGAG CGGGCAGAGC TGATCGCCGC TCTGGAGGCG
49501 CTGGCGACGG GTGAGCCGCA TCGGCGCTG GTCGGCCCGG CTTGCTCGCA
49551 GGCTCGGGTG GGTGGCGATG ACGTGGTGTG GCTGTTCACT GGTGAGGGCA
49601 GTCAGTTGGT CGGTATGGGT GCTGGTTTGT ATGAGCGGTT CCCGGTGTTC
49651 GCGGCTGCGT TTGATGAGGT GTGCGGCCTG TTGGAGGGGC CGTTGGGCGT
49701 GGAGGCGGGT GGGTTGCGGG AGGTGGTGTT CCGTGGCCCG CGGGAGCGGT
49751 TGGATCACAC GGTGTGGGCG CAGGCGGGGT TGTTTGCGCT GCAGGTGGGG
49801 TTGGCCCGGT TGTGGGAGTC GGTGCGGGTG CGGCCGGATG TGGTGCTCGG
49851 GCATTCGATC GGTGAGATCG CGGCCGCGCA TGTGGCGGGG GTTTTTGATC
49901 TGGCGGATGC GTGTCGGGTG GTGGGTGCGC GGGCGCGTTT GATGGGTGGG
49951 CTGCCTGAGG GTGGGGCGAT GTGCGCGGTG CAGGCCACGC CCGCCGAGCT
50001 GGCCGCCGAC GTGGACGGAT CGGCTGTAAG TGTGGCGGCA GTCAACACCC
50051 CCGACTCCAC GGTGATTTTC GGCCCGTCGG ACGAGGTGGA CCGGATTGCT
50101 GGGGTGTGGC GGGAGCGTGG GCGCAAGACG AAGGCGCTGA GCGTCAGTCA
50151 TGCCTTCCAT TCGGCGTTGA TGGAGCCGAT GCTCGCGGAG TTCACCGAAG
50201 CGATACGAGG GGTCAAGTTC AGGCAGCCGT CGATCCCGCT CATGAGCAAT
50251 GTCTCCGGAG AGCGGGCCGG CGAGGAGATC ACGGATCCGG AGTACTGGGC
50301 GAGGCATGTA CGTAATGCGG TGCTCTTCCA GCCCGCCATC GCCCAAGTAG
50351 CGGATTCAGC GGGCGTGTTT GTGGAGCTCG GCGCCGCGCC TGTGCTGACC

50401 ACGGCCGCCC AGCACACCCT GGACGAGTCG GACAGCCAGG AGTCGGTGCT
50451 GGTCGCGTCT CTCGCCGGTG AGCGTCCTGA GGAGTCGGCG TTTGTGGAGG
50501 CGATGGCTCG TCTGCATACC GCTGGTGTTG CTGTGGACTG GTCGGTGTTG
50551 TTCGCGGGTG ATCGTGTGCC TGGGCTGGTG GAGTTGCCGA CGTATGCGTT
50601 CCAGCGGGAG CGGTTCTGGT TGAGTGGCCG TTCTGGGGGT GGGGATGCGG
50651 CGACTTTGGG GTTGGTGGCG GCGGGGCATC CGTTGTTGGG TGCGGCGGTG
50701 GAGTTCGCGG ACCGGGGTGG GTGTCTGCTG ACCGGTCGTC TGTCGCGGTC
50751 TGGGGTGTCG TGGCTTGCTG ATCATGTGGT GGCGGGTGCG GTTTTGGTGC
50801 CGGGTGCTGC GTTGGTGGAG TGGGCGTTGC GGGCCGGTGA TGAGGTCGGT
50851 TGTGTGACGG TGGAGGAGTT GATGTTGCAG GCGCCTTTGG TGGTGCCTGA
50901 GGCCTCGGGT CTGCGGGTTC AGGTGGTGGT TGAGGAGGCG GGTGAGGACG
50951 GGCGGCGCGG TGTTCAATC TACAGCCGGC CCGACGCGGA CGCCGTGGGC
51001 GGCGATGACT CGTGATCTG CCACGCGACC GGCCTACTGT CACCCGAAAG
51051 CGCTCGTCTG GACACGGAGT TGGGTGGCGT CTGGCCACCG GCCGGTGCCG
51101 AACCCTGGA TGTCGACGGC TTCTACGCGC AGGCCGGTGA GGCCGGGTAC
51151 GGATACGGTC CGGCGTTCCG GGGGCTGCGT GCCGTGTGGC GGCACGGCCA
51201 GGACCTGCTG GCCGAGGTCG TCCTGCCCCA AGCCGCCGGT GCCCATGACG
51251 GCTACGGGAT CCACCCCGCC CTCCTCGACG CCACCTCCA TCCGCTGCTC
51301 GCCGCCCGCT TCATGGACGG TTCCGAGGAC GATCAGCTCT ACGTACCGTT
51351 CGGGTGGGCC GGAGTGTCTC TGCGGGCGGT GGGAGCCACG ACTGTGCGCG
51401 TGCGCCTCCG TCCGGTCGGG GAGAGCGTCG ACCAAGGGCT GAGCGTGACG
51451 GTCACCGATG CGACCGGCGG TCCCGTTCTG AGCGTCGACT CCCTCCAGAC
51501 CCGCCCCGTG AAGCCGAGCC AATTGGCTGC GGCCCAACAG CCGGACGTAC
51551 GCGGTCTGTT CACTGTGGAG TGGACGCCGC TGCCGACAGC GGATGCCGAC
51601 GGGGAGGCCG ACTGGGTTGT GCTCTCGGAC GGTGTTGGCC GTCTGGCTGA
51651 TGTGGTGTCG GCGGCGGGTG GTGAAGCGCC GTGGGCAGTG GTCGCTCCTG
51701 TCGATGCGTC TGTGGGCGAC GGCCGTGAGG GTCTTGACGG TCGGCTGGTC
51751 GTGGAGCGGG TGCTGTCACT CGTACAGGAG TTCCTGGCCC TGCCGAGCT

51801 GGCCGAGTCC CGTCTCCTCG TGGTGACGCG CGGTGCGGTG GCCACCGGCG
51851 TCGACGGTGA CGGTGACGTG GACGCGTCCG CCGCAGCTGT ATGGGGCCTG
51901 GTCCGCAGTG CTCAGTCCGA GAATCCGGGC CGCTTCATCC TGCTCGACGT
51951 GGACGGCGAC GGCGACGACC AGGGCCCCGA CCTGAACGGC CGGCATCTGC
52001 CCCACGCCAC CCTGCGTCAC GCCGCCGAGG AACTCGACGA GCCCCAACTC
52051 GCCCTGCGGG AAGGGACGCT CTACGTCCCC CGACTGACCC AGGCGCGCCA
52101 GTCCGCCGAA CTCGTCTGTC CGCCCGGTGA ACCGGCGTGG CGCCTGCGGA
52151 TGGTGACAGA CGGCTCGCTG GACGCCCTGG CGGCAGTGGC CTGCCCCGAG
52201 GCCCTGGAGC CCTTGGCGCC GGGGCAGGTG CGTATCGCCG TACACGCCGC
52251 GGGCATCAAC TTCCGTGACG TACTGGTGGC CTTGGGTATG GTCCCCGCGT
52301 ACGGGGCCAT GGGTGGCGAA GGTGCCGGTG TCGTGACGGA GGTCGGTCCC
52351 GAGGTCACCC ATGTCTCGGT GGGCGACCGC GTGATGGGCG TGTTCGAGGG
52401 CGCGTTCCGC CCTGTGGTGA TCGCCGAGGC GCGGATGGTC ACACCTGTCC
52451 CGCAGGGCTG GGACATGCGG GAGGCGGCCG GTATTCCGGC GGCTTCCTG
52501 ACGGCTTGGT ACGGGTTGGT GGAGCTGGCC GGTCTGAAGG CGGGCGAGCG
52551 GGTGCTGGTC CATGCCGCGA CGGGTGGTGT GGGGATGGCG GCGGTGCAGA
52601 TCGCCCCGCA TGTGGGTGCC GAGGTGTTCG CCACCGCGAG TCCGGGCAAG
52651 CACGCCGTGC TGGAGGAGAT GGGCATCGAC GCCGCCCACC GCGCCTCCTC
52701 CCGGGACCTC GCCTTCGAGG GCACGTTGAG GGAAGCAACG GGCGGCCGCG
52751 GCATGGACGT CGTGCTCAAC AGCCTTGCCG GCGAGTTCAT CGACGCCTCT
52801 CTGCGGTTGC TCGGCGACGG CGGCCGGTTC CTGGAGATGG GCAAGACCGA
52851 TGTGCGGGCC GCCGAAGAGG TGGCTGCGGA GCACGCGGAC GTCTCGTACA
52901 CGGCGTACGA CCTCGTCGGT GATGCCGGAC CCGACCGCAT CAGCAACATG
52951 CTGGACAAGC TCGTCGAATT GTTCGCCTCA GAACGGCTTA AGCCGCTGCC
53001 GGTACGTTCC TGGCCGCTGG ACAAGGCGCA GGAGGCGTTC CGGTTCATGA
53051 GTCAGGCGAA GCACACCGGC AAGCTGGTGC TTGAGATCCC GCCTGCCCTC
53101 GACCCCGAGG GCACGGTTCT GGTCACCGGG GGCACCGGTG CGCTGGGGCA
53151 GGTCGTGGCC GAGCATCTGG TCCGGGAGTG GGGCGTACGG CACCTGCTGC

53201 TGGCCAGCCG TCGCGGTCCG GAGGCGCCGG GCAGCGACGA ACTGGCCTCG
53251 AAGCTCACCG GGTGCGGTGC CGAGGTCACC ATTGTCGCGG CCGATGTCAG
53301 CGACCCGGCC TCGGTGGTGG AGCTGGTCGG CAAGACGGAT CCCTCGCATC
53351 CGTTGACGGG TGTCGTGCAC GCGGCGGGCG TGTGAGGA CGGTGTCGTG
53401 ACCGCTCAGA CGCCTGAGGG GCTGGCGCGG GTGTGGGCGG CCAAGGCTGC
53451 TGCGGCGGCG AATCTCCATG AGGCGACCCG GGAGATGCGT CTCGGCCTGT
53501 TCGTGGTGTT CTCCTCGGCG GCCGCCACGC TCGGCAGTCC GGGCCAGGCC
53551 AACTACGCGG CCGCCAATGC CTATTGCGAC GCGCTGATGC AGCACCGACG
53601 GGCGGTGGGC CAGGTCGGCC TGTCGGTCGG CTGGGTCTC TGGGAGGCGC
53651 CGGACGCCAA GCCGGGTGTT GCCGCCGACG CCAAGGCGAG TGCTGCCACC
53701 GTCGGCAAGG CGAGTGCTCT ATCCGACGGC ACGAACGGCA GCGCTCCCCA
53751 GGACACGACC GGCACCGCCC CCCAGGGCAT GACCGGCGGA CTCACCGACA
53801 CCGACGTAGC CCGCATGGCA CGTATCGGCG TCAAGGGCAT GAGCAACGCC
53851 CACGGTCTCG CCCTGTTCGA CGCCGCGCAC CGCCACGGCC GCCCCACCT
53901 GGTCGGCTTC AACCTCGACC TGCGCACCCCT GGCCACGCAC CCCCTGCACA
53951 CCCGGCCCCG CTTCTGCGC GGCCTGGCCA CCCCCACCGC CGGCGGGGCG
54001 AGCAGGCCGA CCGCGACGGC GGGCGGACAG CCCGCCGACC TGGCGGGCCG
54051 GCTGGCCGCG CTGTGCGCGT CGGACCGGCA CCACACGCTG GTCCGGCTCA
54101 TCAGGGAACA GGCCGCCACC GTGCTCGGGC ACCACCCGGA CAGTCTCACC
54151 ACGGGCAGCA CCTTCAAGGA ACTCGGATTC GACTCCCTGA CCGCGGTCTGA
54201 ACTGCGCAAC AGGCTGTCCG CCGCCACCGG TCTCCGGCTC CCCGCCGGCC
54251 TGGTCTTCGA CCACCCGGAC GCCGACATCC TGGCCGAACA CCTCGGCGCG
54301 CAACTCGCCC CCGACGGGGA CACCCCGGCC GGTGCGGAAG CCACCGACCC
54351 GGTCTCTCGC GACCTGGCGA AACTCGAGAA CGCCCTCTCC TCCACCCTCG
54401 TCGAGCACCT CGACGCCGAC GCGGTCACGG CCCGACTGGA AGCACTCCTG
54451 TCGAACTGGA AGGCGGCGAG CGCGGCGCCC GGCTCGGGCA GCACGAAGGA
54501 GCAGCTCCAG GTTGCCACGA CCGACCAGGT CCTCGACTTC ATCGACAAAG
54551 AACTGGGTGT GTGAAACGAC CGTGCACGGC GCGACAACCA CGCTGAAGGC

54601 TGGGTGAACT CTCATGGCGA GTGAAGAGGA ACTGGTCGAC TACCTCAAGC
54651 GGGTCGCCGC CGAACTGCAC GACACCCGGC AGCGCCTGCG CGAGGTCGAG
54701 GACCGGCGGC AGGAGCCGGT GGCCGTCGTC GGCATGGCCT GCCGTTTCCC
54751 CGGCGGCATC GAGACGCCCG AGGGACTGTG GGAGCTGGTC GCGGCCGGCG
54801 ACGACGCCAT TGAGCCCTTC CCCACCGACC GGGGCTGGGA CCTGGAAGGC
54851 ATCTACCACC CGGACCCCGA CCACCCGGGT ACCTGCTACG TCCGGGAGGG
54901 CGGGTTCCTA GCCGCCCTTG ACCGGTTCGA CTCCGACTTC TTCGGCTTCA
54951 GCGCGCGCGA GGCCCTGGCC AGCAGCCCGC AACTGCGACT GCTCCTGGAG
55001 ACGTCCTGGG AGGCCCTCGA ACGGGCGGGC ATCAACCCCG CCTCGCTCAA
55051 GGGCAGCCCC ACCGGCGTCT ACGTCGGCGC CGCGACCACC GGCAACCAGA
55101 CGCAGGGCGA CCCC GGCGGC AAGGCGACCG AGGGTTACGC GGGCACCGCG
55151 CCCAGCGTCC TCTCGGGCCG CCTCTCGTTC ACGCTCGGCC TGGAGGGCCC
55201 GGCGGTGACC GTCGAGACAG CGTGCTCCTC CTCGCTGGTG GCGATGCACC
55251 TGGCGGCCAA CGCCCTGCGC CAGGGCGAGT GCGACCTCGC CCTCGCGGGC
55301 GGCGTCACCG TCATGTCCAC CCCCAGAGTG TTCACAGGCT TCTCGCGTCA
55351 GCGGGGACTG GCCCCGACG GCCGCTGCAA GCCGTTCGCC GCCGCGGCCG
55401 ACGGCACGGG CTGGGGCGAG GGCGCGGGCC TGATCCTCCT GGAGCGCCTC
55451 TCCGACGCCC GCAGGAAGGG CCACAAGGTC CTCGCGGTGA TCCGGGGCTC
55501 GGCGATCAAC CAGGACGGCG CGAGCAACGG CTTACCGCG CCCAACGGCC
55551 CCTCGCAGCG CCGCGTCATC CGCCAGGCAC TCTCCAGCGC CCACCTCTCC
55601 ACGTCGGAGA TCGACGTCGT CGAGGCGCAC GGCACCGGCA CCAGGCTCGG
55651 CGACCCCATC GAGGCCGAGG CGCTCATCGC CACCTACGGC AAGGAGCGCG
55701 AGGACGACCG TCCCCTGTGG CTCGGCTCGG TCAAGTCCAA CATCGGCCAC
55751 ACGCAGGCCG CCGCGGGCGT CGCCGGAGTC ATCAAGATGG TGATGGCGCT
55801 ACAGCGCGAA CTGCTTCCCG CCACCCTGAA CGTCGACGAG CCGACCCCGC
55851 ACGTCCAGTG GGAGGGCGGC GGCGTACGCC TCCTGACCGA ACCGGTCCCCG
55901 TGGTCGCGCG GCGAACGCCC GCGCCGCGCC GGAATCTCCT CCTTCGGCAT
55951 ATCGGGCACG AACGCGCACG TGGTCCTGGA GGAGGCGCCG CCGGAGGAGG

56001 ACGTGCCGGG CCCCGTGGCT GCGGAGCCGG AAGGGGTGGT GCCGTGGGTG
56051 GTCTCCGCGC GGACCGAGGA GGC GTTGAGC GAACAGGCGC GGCGCCTGGG
56101 CGAGTTCGTG GCCGACACGG ACCCGTCGAC CGCTGACGTC GGGTGGTCAC
56151 TGACCACGAG CAGGGCGATC CTTGAACACC GCGCTGTGGT GGTGGGGCGT
56201 GATCGGGATG CGCTGACGGC CGGCCTGGCG GCGTTGGCCG CGGGTGAGGA
56251 GTCGGCGGAT GTGGTGGCTG GGGTGGCCCG TGATGTGGGT CCTGGGCCGG
56301 TGT TGGTGT TCCGGGGCAG GGGTCGCAGT GGGTGGGCAT GGGCGCCCAG
56351 CTCCTTGACG AGTCGCCCCGT CTTGCGGGCG CGGATCGCGG AGTGTGAGCA
56401 GGCCTGTGCG GCGTACGTGG ACTGGTCGCT GAGTGCGGTG TTGCGCGGGG
56451 ATGGGAGTGA ACTGTCCCGG GTCGAGGTCG TGCAGCCGGT GTTGTGGGCG
56501 GTGATGGTCT CGCTGGCTGC CGTCTGGGCG GATTACGGGG TCACCCCGGC
56551 CGCTGTGATC GGGCACTCGC AGGGCGAGAT GGCCGCCGCG TCGTGGCGG
56601 GGGCGCTGTC TTTGGAGGAT GCGGCGCGCG TCGTGGCCGT ACGCAGTGAC
56651 GCGCTTCGTC AGCTGATGGG GCAGGGCGAC ATGGCGTCGT TGGGCGCCAG
56701 CTCGGAGCAG GCGGCTGAGC TCATCGGTGA TCGGCCGGGC GTATGCATCG
56751 CAGCGGTCAA CGGGCCGTCC TCGACAGTCA TTTCAGGACC GCCGGAGCAT
56801 GTGGCAGCCG TGGTCGCGGA TCGGAGGAA CGTGGTCTGC GCGCCCGTGT
56851 CATCGATGTC GGCTATGCCT CGCACGGTCC CCAGATCGAT CAGCTCCACG
56901 ACCTCCTCAC CGACCGGCTC GCCGACATCC GGCCCGCGAC CACGGACGTG
56951 GCCTTCTATT CGACGGTCAC CGCCGAGCGC CTGACGGACA CCACGGCCCT
57001 GGATACGGAT TACTGGGTTA CCAACCTCCG CCAGCCGGTC CGTTTCGCCG
57051 ACACCATCGA TGCGCTTCTC GCGGACGGCT ATCGCCTGTT CATCGAGGCC
57101 AGCGCGCACC CGGTGCTGGG TCTGGGCATG GAGGAGACCA TCGAGCAGGC
57151 GGACATCCCC GCCACGGTCG TCCCCACCCT GCGCCGCGAT CACGGTGACA
57201 CCACCCAGCT CACCCGTGCC GCAGCGCACG CCTTCACCGC CGGCGCCACC
57251 GTCGACTGGC GGCCTGGTT CCCGGCCGAC CCCACCCCCC GCACGATCGA
57301 CCTGCCCACC TACGCCTTCC AGCGCCGCAG CTA CTGGTTG CCGGTGGACG
57351 GTGTCGGAGA TGTGCGGTG GCCGGGCTGC GGCGGGTGGA AACTCGCTG

57401 TTGCCC GCGG CGCTCGGTCT CGCCGATGGT GCGCTCGTGC TGACCGGACG
57451 GCTCGCGGCG TCCGGTGGTG GTGGCGGTTG GCTCGCGGAT CACGCGGTGG
57501 CGGGCACGAC GCTCGTCCCC GGTGCCGCGC TGGTCGAGTG GGCGTTGCGG
57551 GCCGCCGACG AGGCGGGCTG CCCCTCCCTT GAGGAGCTGA CGCTCCAGGC
57601 ACCTCTGGTG CTGCCCCGGCT CCGGGGGCCT CCAGGTCCAA GTGGTCTGTG
57651 GTCCGGCCGA CGGACAGGGC GGCCGGCGTG AGGTGCGCGT CTTCTCGCGT
57701 GTCGACTCGG ACGACGAGGC AGCGGGGCAG GACGAGGGGT GTTCGTGTCA
57751 CGCGACCGGT GTGCTGAGCC CCGAGCCCGG TCGGGTACCG GACGGGCTCA
57801 GCGGACAGTG GCCCCGACG GGCGCCGAGC CGCTGGAGAT CAGTGATCTC
57851 TACGAGCAGG CGGCATCGGC GGGATACGAG TACGGGCCGT CGTTCCGGGG
57901 CCTGCGCTCC GTGTGGCGGC ACGGGCATAA CCTGCTGGCA GAGGTGGAGC
57951 TGCCCGAACA GGCAGGTGCG CACGACGACT TCGGCATCCA CCCCCTACTG
58001 CTGGACGCCG CGCTGCACCC GGCGCTGCTG CTCGACCAGA ACGCGCCCGG
58051 CGAAGAGCAA GAGCCAGCCC AGCCCCTCTT TCGCCTGCCG TTCGTGTGGA
58101 ACGGCGTCTC CCTGTGGGCC ACCGGCGCCG CGACCGTGCG GGTACGGCTG
58151 GCCCCGCACG GGGGAGGGGA GACGGACGAT AGCGCCGGGC TGCGCGTGAC
58201 GGTGCGCGAC GCCACCGGAG CACCGGTGCT GAGCGTGGAC TCCCTCGCTC
58251 TGCGCCCCGC TGACCCCGAA CTGCTGCGCA CGGCCGGTCG GGCGGGCAGC
58301 GGCACCAACG GCTTGTTTAC GGTGGAGTGG ACCGCTCTGC CCCC GGCGGA
58351 CGTGGCCGAC CACGCCGCAG GCGACGGCTG GGCGGTGCTC GGTCAGGACG
58401 TACCCGACTG GGCCGGAGCG GACATGCCCC GGCATCCCGA CATGGCCTCC
58451 CTGTGCGCCG CGCTGGACGA GGGAACGCAG GCCCCTGCGG CCGTCTTCGT
58501 GGAGACCACA GCCACATCGC ACGCCACACC GAACACCGCA GCGGACGTGA
58551 CGCTCGACGC GTCCGGCCCG GCGGTCGCCG AGCGCACCTT GCACCTGCTG
58601 CGGGACTGGC TCGCCGAACC GCGCCTCGCC GAGACCCGGC TCGTCCTCAT
58651 CACCCACCAC GCGGTGACGA CCCC GGCGGA CGACGACGTG AACGCCGCAC
58701 CCCTCGACGT CCCGGCCGCC GCCCTGTGGG GACTGATCCG CAGCGCACAG
58751 GCCGAACACC CGGACCGCTT CGTTCTGTTG GACACCGACG CGAAGGCCAA

58801 CACCGACCCC GGCCCCGACA CCAGTACTGA CCACAGCACC GCATCGGGTA
58851 CGTACCGAAC CGTCATCGCG CGGGCCCTCG CCACCGGGGA GCCACAGCTG
58901 GCCGTGCGCG CGGGAGAACT GCTGGCTCCC CGCCTCGCCC GAGCCGCCAC
58951 CCCCACACCC GAGACCCCCA CACCCGAGAC ACAGCCCGAC ACCGGATCCG
59001 GGTCCGAGGC CGGGGCCGGG TCCGGATCTG GACCCGGCGC GACACTGGAC
59051 CCCGACGGCA CCGTCCTCAT CGCGGGCGGC ACCGGCATGA TGGGTGGTCT
59101 CGTCGCCGAA CACCTGGTCC GCGCCTGGTC GGTGCGGCAC CTCCTGCTCG
59151 TCAGCCGGCA AGGGCCCGAC GCGCCGGACG CCCGCGACCT CGCCGACCGG
59201 CTGGTCGGCC TGGGCGCGAC GGTACGGATC GTCGCGGCCG ACCTGACGGA
59251 CGGGCGGGCC ACCGCGGACC TCGTCGCGTC GGTGACCCG GCGCACCCGC
59301 TCACCGGTGT GATCCACGCG GCCGGCGTCC TGGACGACGC CGTGGTCACC
59351 GCGCAGACCT CCGACCAGCT GGCCAGGGTG TGGGCGGCCA AGGCGTCCGT
59401 CGCCGCCAAC CTGGACGCGG CCACGTCGGA GCTGCCGCTC GGCTTGTTCC
59451 TGATGTTCTC GTCCGCCGCC GGTGTCCTCG GCAACGCGGG CCAGGCCGGT
59501 TACGCGGCCG CCAACGCCTT CGTCGACGCC CTGGTCGGCC GCCGTCGCGC
59551 CACCGGCCTG CCCGGCCTGT CGATCGCCTG GGGCCTGTGG GCGCGCGGCA
59601 GCGCCATGAC CCGGCACCTG GACGACGCCG ACCTCGCGCG GCTGCGTGCC
59651 GGCGGGGTCA AGCCCCTGCT GGACGAGCAG GGCCTCGCCC TCCTCGACGC
59701 GGCGCGCGCC ACCGCCGCGC ACACCTCGCT GGTGGTCGCG GCCGGTATCG
59751 ACGTACGCGG ACTGAACAGG GACGACGTCC CCGCGATCCT CCGCGACCTG
59801 GCGGGCCGGA CCCGCCGAG GCGGGCCGCC GACTCCACCG TCGACCAGGC
59851 CGCGCTGGAG CGGCGCCTCA CGGGCCTGGA CGAGGCCGAG CGCCGGGCTG
59901 TCGTCACCGA CGTCGTACGC GAATGCGTGG CGGCCGTGCT CGGCCACCGG
59951 TCGGCGGCCG ACGTACGCAC CGAGGCCAAC TTCAAGGACC TCGGCTTCGA
60001 CTCGCTCACT GCGGTGCAGC TGCGCAACCG CCTCTCGGCG GCGAGCGGCC
60051 TCCGCCTGCC CGCCACCCTG GCCTTCGACC ACCCCACCCC CCAGGCGCTG
60101 GCGGCGTACC TCGGCACGCG CCTGAGCGGC CGGACCGCCA CCCCCTCGC
60151 ACCCGTGGCG CCTTCCGCGG CCGCGACGGA CGAGCCGGTG GCGATCGTCG

60201 CGATGGCCTG CAAGTACCCG GGTGGAGCGA CCTCGCCGGA AGGCCTCTGG
60251 GACCTGGTCG CGGAGGGCGT GGACGCGGTC GGCGCCTTCC CGACGGGCCC
60301 CGGCTGGGAC CTCGAACGGC TCTTCCACCC CGACCCGGAC CACCCCGGCA
60351 CGAGTTACGC CGACGAAGGG GCCTTCCTTC CTGACGCGGG CGATTTCGAT
60401 GCGGCGTTCT TCGGGATCAA TCCGCGGGAG GCGCTGGCGA TGGATCCGCA
60451 GCAGCGGCTG TTGCTGGAGG CGTCGTGGGA GGTGTTGGAG CGTGCGGGTA
60501 TCGACCCGAC GACGCTCAAG GGCACCCCGA CCGGCACGTA CGTCGGCGTG
60551 ATGTACCACG ACTACGCGGC AGGCCTCGCC CAGGACGCCC AACTGGAGGG
60601 CTA CTCCATG CTCGCCGGCT CCGGCAGCGT GGTGTCCGGC CGCGTCGCCT
60651 ACACCCTGGG GCTTGAGGGT CCTGCGGTGA CGGTCGACAC CGCGTGCTCC
60701 TCGTCCCTGG TCTCCATCCA CCTGGCCGCG CAAGCACTGC GACAGGGCGA
60751 GTGCACTCTC GCCCTCGCGG GCGGCGTGAC CGTCATGGCC ACGCCCGAGG
60801 TGTTACCGG ATTCTCGCGC CAGCGCGGCC TGGCCCCCGA CGGCCGCTGC
60851 AAGCCGTTCTG CCGCCGCCGC CGACGGCACC GGCTGGGGCG AGGGTGTCTGG
60901 TGTGTTGTTG CTCGAGCGGT TGTCGGATGC GCGGCGTCAT GGGCGTCGGG
60951 TGTGGGTGT GGTGCGGGGT TCGGCGGTGA ATCAGGACGG TCGAGTAAT
61001 GGGTTGACGG CGCCGAATGG TCCGTGCGAG GAGCGGGTGA TTCGTGAGGC
61051 GTTGGCCAGT GGTGGGTTGT CGTCGGTGGA TGTGATGTG GTGGAGGGGC
61101 ATGGGACGGG GACCACGTTG GGTGATCCGA TCGAGGCGCA GGCTCTGCTG
61151 GCCACGTATG GGCAGGGGCG TCCGGTGGAT CGTCCGTTGT GGTTGGGGTC
61201 GGTGAAGTCG AATATTGGTC ATACGCAGGC GGCTGCGGGT GTTGCGGGTG
61251 TCATCAAGAT GGTGATGGCG ATGCGGCATG GTGTGGTGCC GGCAGTTTG
61301 CATGTGGATG TGCCGTCGCC GCATGTGGAG TGGGATTCCG GTGCGGTGCG
61351 GTTGGCGGTT GAGTCGGTGC CATGGCCGGA GGTGGAGGGT CGTCCGCGTC
61401 GGGCGGGTGT GTCGTCGTTT GGGGCTTCGG GAACGAATGC GCACGTGATC
61451 GTGGAGTCTG TGCCCGATGG GCTGGGGGAG GACTCGGTAT CGGTGAGTGG
61501 TGAGGCTCCC GAGACTGAGA CTGACGGGCG CTTGGTGCCG TGGGTGGTAT
61551 CGGCCCCGAG CCCGCAGGCC CTGCGCGACC AGGCACTACG CCTGCGTGAT

61601 GCGGTGGCGG CCGACTCAAC GGTGTCGGTG CAGGATGTGG GCTGGTCGCT
61651 GCTGAAGACG CGTGCGCTGT TCGAGCAGCG GCGGTGGTG GTGGGGCGTG
61701 AGAGGGCTGA ACTCCTGTCG GGGCTTGCTG TGTGGCCGC TGGCGAGGAG
61751 CACCCGGCTG TGACGCGGTC CCGTGAGGAC GGGGTTGCTG CGAGCGGTGC
61801 TGTGGTGTGG CTGTTCACTG GTCAGGGCAG TCAGTTGGTC GGTATGGGTG
61851 CTGGTTTGTA TGAGCGGTTT CCGGTGTTTG CGGCTGCGTT TGATGAGGTG
61901 TGCGGCCTGT TGGAGGGGCC GTTGGGCGTG GAGGCGGGTG GGTGCGGGA
61951 GGTGGTGTTC CGTGGCCCGA GGGAGCGGTT GGATCACACG ATGTGGGCGC
62001 AGGCGGGGTT GTTTGCGCTG CAGGTGGGGT TGGCCCGGTT GTGGGAGTCG
62051 GTCGGGGTGC GGCCGGATGT GGTGCTCGGG CATTCGATCG GTGAGATCGC
62101 GGCCGCGCAT GTGGCGGGGG TCTTTGATCT GCGGATGCC TGTCGGGTGG
62151 TGGGGGCGCG GGCCCGTTTG ATGGGTGGGC TGCCTGAGGG CGGGGCGATG
62201 TGCGCGGTGC AGGCCACGCC CGCCGAGCTG GCCGCCGACG TGGACGACTC
62251 TGGTGTGAGT GTGGCGGCGG TCAACACACC TGATTCGACG GTGATTTTCA
62301 GGCCGTCTGG TGAGGTGGAT CGGATTGCTG GGGTGTGGCG GGAGCGTGGG
62351 CGTAAGACGA AGGCGCTGAG CGTCAGTCAT GCCTTCCACT CGGCGTTGAT
62401 GGAGCCGATG CTCGCGGAGT TCACCGAAGC GATACGAGAG GTCAAGTTCA
62451 CGCGGCCGAA GGTGTCGTTG ATCAGCAACG TCTCTGGTCT GGAGGCGGGT
62501 GAGGAGATCG CGTCCCCGGA GTACTGGGCA CGCCATGTAC GCCAGACAGT
62551 GCTCTTCCAG CCCGGCATCG CCCAAGTGGC TTCCACGGCA GCGTGTTTG
62601 TCGAGCTCGG CCCCGGCCCC GTACTGACTA CTGCCGCCCA GCACACCCTG
62651 GACGACGTAA CCGATAGGCA TGGCCCCGAA CCGGTACTGG TGTCTCGCT
62701 GGCCGGTGAG CGTCCTGAGG AGTCGGCGTT CGTGGAGGCG ATGGCTCGTC
62751 TGCATACCGC TGGTGTGCT GTGGACTGGT CGGTGTGTT CGCGGGTGAT
62801 CGTGTCCTG GGCTGGTGGA GTTGCCGACG TATGCGTTCC AGCGGGAGCG
62851 GTTCTGGTTG AGCGGCCGTT CTGGGGGTGG GGATGCGGCG ACTTTGGGTC
62901 TGGTGGCGGC GGGGCATCCG TTGTTGGGTG CGGCGGTGGA GTTCGCGGAC
62951 CGGGGTGGGT GTCTGCTGAC CGGTCGGCTG TCGCGGTCTG GGGTGTCTG

63001 GCTTGCTGAT CATGTGGTGG CGGGTGCGGT TTTGGTGCCG GGTGCTGCGT
63051 TGGTGGAGTG GGC GTTGCGG GCCGGTGATG AGGTCGGTTG TGTGACGGTG
63101 GAGGAGTTGA TGTTGCAGGC GCCTTTGGTG GTGCCTGAGG CGTCGGGTCT
63151 GCGGGTTCAG GTGGTGGTCG AGGAGGCGGG TGAGGACGGG CGGCGCGGTG
63201 TCCAGATCTA TAGCCGGCCT GACGCGGACG CCGTGAGCGG CGACGACTCG
63251 TGGATCTGCC ACGCGACCGG CACCCTCACC CCCCAGCACA CCGACGCTCC
63301 GAACGACGGA CTGGCCGGCG CGTGGCCCCG GCGGGGCGCC GTGCCGGTGG
63351 ACCTGGCGGG CTTCTACGAG CGCGTGGCGG ACGCGGGCTA TGC GTACGGC
63401 CCGGGGTTC AGGGGCTGCG TGCCGTGTGG CGGCACGGTC AGGACCTGCT
63451 GGCCGAGGTC GTCCTGCCCC AAGCCGCGGG TGCCCATGAC GGCTACGGCA
63501 TCCACCCCGC CCTCCTCGAC GCCACCCTCC ACCCGGCCCT GCTCCTCGAC
63551 TGGCCCGGG AGGTGCAGGA CGACGACGGG AAGGTCTGGC TGCCTTTTAC
63601 CTGGAACCAG GTCTCCTTGC GGGCTGCGGG AGCCGCCACC GTACGCGTAC
63651 GTCTCTCGCC CGGCGAGCAC GACGAGGCGG AACGGGAAGT ACAGGTACTG
63701 GTGGCCGACG CCACCGGGAC CGACGTCCTG AGCGTGGGGT CGGTGACGTT
63751 GCGTCCCGCC GACATCCGGC AACTGCAGGC CGTGCCGGGT CACGACGACG
63801 GTCTGTTCTC GGTGGACTGG ACGCCGCTGC CGCTGTCGCG GACGGATGTG
63851 TCGCAGACGG ATGCCGACGG GGATGCCGAC TGGGTTGTGC TCTCGGACGG
63901 TGTCGGCAGC CTGGCTGATG TGGTGTGCGC GCGGGGTGGT GAAGCGCCGT
63951 GGGCAGTGGT CGCTCCCGTC GGTGCATCCG CGGGCGGCGG CCTTGCCGGC
64001 TTTGACCGCC GTGAGGGTCT TGACGGTCGG CTGGTCGTGG AGCGGGTGTT
64051 GTCACTCGTA CAGGAGTTCC TGGCCGCGCC GGAGCTGGCC GAGTCCCGGC
64101 TCCTCGTGCT GACCCGCGGC GCCGTGGCGA CCGGCGGCGA CGGCGACGGT
64151 GATGTGGACG CGTCCGCCGC AGCCGTATGG GGCCTGGTCC GCAGTGCTCA
64201 GTCCGAGAAC CCGGGCCGCT TCATCCTGCT CGACGTGGAC ATGGACGTGG
64251 ACGTCGACGT GGACATGGAC GTGGACGTCG ACGTGGACGT CGACGTGGAC
64301 GTGGACGGAG ACGGCAATGG CAGCGACCTG GACCCGGACC TGAACGGCCC
64351 ACGACTTCCC CACGCCACCC TGC GTACGC CGCCGAGGAA CTCGACGAGC

64401 CCCAACTCGC CCTGCGCGAC GGACAACTGC TCGTTCCGCG GCTGGTCCGC
64451 GCCACCGGCG GCGGACTCGT CGTGGCGCCC ACCGACCGTG CCTGGCGCCT
64501 GGACAAGGGA AGCGCCGAGA CGCTGGAGAG CGTCGCGCCG GTCGCGTACC
64551 CCGGAGTCAT GGAACCCCTG GGCCCCGGCC AGGTCCGCCT CGGCATCCAC
64601 GCCGCGGGCA TCAACTTCCG CGACGTCCTG GTCAGCCTCG GCATGGTGCC
64651 CGGCCAGGTC GGCCTGGGCG GCGAAGGCGC CGGTGTCGTG ACGGAGACAG
64701 GCCCCGATGT CACCCACCTG TCGGTCGGCG ACCGCGTGAT GGGCGTCCTC
64751 CACGGCTCCT TCGGCCCGAC GGCCGTGGCG GACACCCGCA TGGTCGCGCC
64801 GGTTCGCGAG GGCTGGGACA TGCGGCAGGC GGCCGCGATG CCCGTCGCGT
64851 ATCTGACGGC TTGGTACGGG TTGGTGGAGC TGGCCGGTCT GAAGGCGGGC
64901 GAGCGCGTGC TGATCCACGC AGCCACGGGT GGTGTGGGAA TGGCGGCGGT
64951 GCAGATCGCC CGTCACCTGG GTGCCGAGGT GTTCGCCACC GCCAGTGCAG
65001 CCAAGCACGT CGTACTGGAA GAGATGGGCA TCGACGCCGC CCACCGCGCC
65051 TCCTCCCGGG ACCTCGCCTT CGAGGACACC TTCCGGCAGG CCACCGACGG
65101 GCGCGGCATG GACGTCGTCC TCAACAGCCT GACCGGCGAG TTCATCGACG
65151 CATCTCTGCG GTTGCTCGGC GACGGCGGCC GGTTCCTGGA GATGGGCAAG
65201 ACCGATGTGC GCACGCCGGA GGAGGTGGCC GCGGAGTACC CGGGTGTCAC
65251 CTACACCGTG TACGACCTCG TCACCGACGC GGGGCCGGAT CGCATCGCGG
65301 TCATGATGAG TGAGCTGGGC GAGAGGTTG CTTCCGGTGC CTTGACCCT
65351 CTGCCGGTGC GTTCCTGGCC GCTGGACAAG GCGCGTGAGG CGTTCCGGTT
65401 CATGAGTCAG GCCAAGCACA CCGGCAAAC TCGTACTCGAC GTGCCCCGAC
65451 CGCTCGACCC CGACGGGACC GTCCTGATCA CCGGAGGCAC GGGGGCGCTG
65501 GGGCAGGTCG TGGCCGAGCA TCTGGTGC GGAGTGGGGCG TACGGCACCT
65551 GCTGCTGGCC AGCCGCCGTG GACTGGACGC CCCCAGGAGC GGTGAACTCG
65601 CCGACAGGCT GTCGGAATTG GGCGCCGAGG TGACCGTCGC GGCGGCCGAT
65651 GTGAGCGACC CGGCCTCGGT GGTGGAGCTG GTCGGCAAGA CGGATCCCTC
65701 GCATCCGTTG ACGGGTGTCG TGCACGCGGC GGGCGTGCTT GAGGACGGGA
65751 TCGTGACGGC TCAGACGCCT GAGGGGCTGG CGCGGGTGTG GGCGGCCAAG

65801 GCCGCTGCGG CGGCGAATCT CCATGAGGCG ACCCGGGAGA TGCGTCTCGG
65851 TCTGTTTCGTG GTGTTCTCCT CGGCGGCCGC CACGCTCGGC AGTCCGGGCC
65901 AGGCCAACTA CGCGGCTGCC AATGCCTATT GTGACGCGCT GATGCAGCGC
65951 CGACGGGCGG CGGGCCAGGT CGGCCTGTCG GTCGGCTGGG GTCTCTGGGA
66001 GGCACCGGAC GCCAAGCCGG GTGTTGCCGC CGACGCCAAA CCGGATGTTG
66051 CCGCCGACGC CAAGACGGGA GTTGCCGCCG ACGGCACTCC CCAGGGCATG
66101 ACCGGCACCC TGAGCGGCAC CGACGTGGCC CGCATGGCAC GCATCGGCGT
66151 CAAGGCGATG ACCAGCGCAC ACGGTCTCGC CCTGCTCGAC GCCGCACACC
66201 GCCACGGCCG CCCCCACCTC GTCGCCGTCG ACCTCGACAC CCGCGTCCTG
66251 GCGCACAAAC CCGCCCCGGC CCTCCCCGCC CTCCTGCGCG CCTTCGCCGG
66301 AGACCAGGGA GGCCAGGGAG GCGGCCGAGG CGGCGGTCGG GGCGGCGGCC
66351 CGGCACGACC GGCGGCCGCC ACCACCCGGC AGAACGTCGA CTGGGCCGCG
66401 AAGCTCTCCG TCCTGACAGC CGAGGAACAG CACCGCACCC TCCTCGACCT
66451 GGTACGGACG CACGCGGCAG CCGTCCTCGG GCACGCGGGC ACCGACGCCG
66501 TACGCGCCGA CGCCGCCTTC CAGGATCTCG GCTTCGACTC CCTCACCGCG
66551 GTCGAACTGC GCAACCGCCT CTCCGCCTCC ACCGGCCTGC GCCTGCCCCG
66601 CACGTTTCATC TTCCGGCACC CGACCCCGTC GGCCATCGCC GACGAACTGC
66651 GCGCACAGCT GGCCCCCGCG GGGGCCGACC CGGCCGCGCC GCTCTTCGGT
66701 GAACTGGACA AGCTGGAGAC GGTGATCACG GGGCACGCGC ACGACGAGAG
66751 CACCCGGACC CGCCTGGCGG CACGCCTGCA GAACCTGCTG TGGCGCCTGG
66801 ACGACACTTC GGCCCGCTCG GACCACGCGG CCGGCGCGAG CGACGCCGAC
66851 GGCGACGCCG TCGAGAACCG AGACCTCGAG TCCGCGTCGG ACGACGAGCT
66901 CTTCGAGCTG ATCGACCGAG AACTGCCTTC TTGATCAGGA GTGGAGAAGA
66951 CATGCCGGGT ACGAACGACA TGCCGGGTAC CGAGGACAAG CTCCGCCACT
67001 ACCTGAAGCG AGTGACCGCG GATCTCGGAC AGACCCGTCA GCGCCTGCGC
67051 GACGTGGAGG AGCGCCAGCG GGAACCGATC GCCATCGTCG CGATGGCCTG
67101 CCGCTACCCG GGCGGGGTGG CCTCCCCCGA GCAGCTGTGG GACCTGGTGC
67151 CCTCACGCGG CGACGCCATC GAGGAGTTCC CCGCCGACCG CGGCTGGGAC

67201 GTGGCGGGCC TCTACCACCC CGACCCGGAC CACCCCGGCA CGACCTATGT
67251 ACGAGAGGCC GGATTCTGTC GGGACGCCGC CCGCTTCGAC GCCGACTTCT
67301 TCGGCATCAA CCCGCGCGAG GCGCTCGCCG CCGACCCGCA GCAACGGGTG
67351 CTCCTCGAAG TGTCGTGGGA ACTGTTTCGAG CGGGCGGGCA TCGACCCCGC
67401 CACGCTCAAG GACACCCTCA CCGGCGTGTA CGCGGGGGTG TCCAGCCAGG
67451 ACCACATGTC CGGGAGCCGG GTCCCGCCGG AGGTCGAGGG CTACGCCACC
67501 ACGGGAACCC TCTCCAGCGT CATCTCCGGC CGCATCGCCT ACACCTTCGG
67551 CCTGGAGGGC CCGGCGGTGA CGCTCGACAC GGCCTGCTCG GCATCGCTGG
67601 TCGCGATCCA CCTCGCCTGC CAGGCCCTGC GCCAGGGCGA CTGCGGCCTG
67651 GCGGTGGCGG GAGGCGTGAC CGTACTGTCC ACGCCGACGG CGTTCGTGGA
67701 GTTCTCACGC CAGCGCGGAC TCGCACCGGA CGGCCGCTGC AAGCCGTTTCG
67751 CCGAGGCCGC CGACGGCACC GGATTCTCCG AGGGCGTCGG CCTGATCCTC
67801 CTGGAACGCC TCTCCGACGC CCGCCGCAAC GGACATCAAG TACTCGGCGT
67851 CGTACGCGGA TCGGCCGTCA ACCAGGACGG CGCGAGCAAC GGCCTGACCG
67901 CCCCGAACGA CGTCGCCCAG GAACGCGTGA TCCGCCAGGC CCTGACCAAC
67951 GCCCGCGTCA CCCCAGACGC CGTCGACGCC GTGGAGGCAC ACGGCACCGG
68001 CACCACGCTC GGCGACCCGA TCGAGGGGAA CGCACTCCTC GCGACGTACG
68051 GAAAGGACCG CCCC GCCGAC CGGCCGCTGT GGCTCGGCTC TGTGAAGTCG
68101 AACATCGGCC ACACGCAGGC GGCTGCGGGC GTCGCAGGCG TCATCAAGAT
68151 GGTGATGGCG ATGCGCCACG GCGAGCTGCC CGCCTCCCTG CACATCGACC
68201 GGCCACGCC CCACGTGGAC TGGGAGGGCG GGGGAGTGCG GTTGCTCACC
68251 GATCCCGTGC CGTGGCCACG GGCCGACCGC CCCC GCCGCG CGGGGGTCTC
68301 CTCCTTCGGC ATCAGCGGCA CCAACGCCCA CCTGATCGTG GAACAGGCCC
68351 CCGCCCCGCC CGACACGGCC GACGACGCC CGGAAGGCGC CGCAACCCCC
68401 GGCCTTCCG ACGGCCTCGT GGTGCCGTGG GTGGTGTGCG CCCGTAGTCC
68451 GCAGGCCCTG CGTGATCAGG CCCTGCGTCT GCGCGACTTT GCCGGTGACG
68501 CGTCCCAGC GCCGCTCACC GACGTGGGCT GGTCTTTGCT GCGGTCGCGT
68551 GCGCTGTTCG AGCAGCGGGC GGTGGTGGCG GGGCGTGAGA GGGCTGAACT

68601 GCTGGCGGGG CTGGCTGCGT TGGCCGCTGG TGAGGAGCAC CCGGCTGTGA
68651 CGCGGTCCCC TGAGGAAGCG GCGGTTGCTG CGAGCGGTGA TGTGGTGTGG
68701 CTGTTCAAGT GTCAGGGCAG TCAGTTGGTC GGTATGGGTG CTGGTTTGTG
68751 TGAGCGGTTC CCGGTGTTTG CGGCTGCGTT TGATGAGGTG TCGGGCTTGC
68801 TGGAGGGGGA GCTGGGGGTT GGTTGCGGTG GGTGCGGGA GGTGGTGTTC
68851 TGGGGCCCCG GGGAGCGGTT GGATCACACG GTGTGGGCGC AGGCGGGGTT
68901 GTTTGCGTTG CAGGTGGGGT TGGCCCGGTT GTGGGAGTCG GTCGGGGTGC
68951 GGCCGGATGT GGTGCTCGGG CATTGATCG GTGAGATCGC GGCCGCGCAT
69001 GTGGCGGGGG TCTTTGATCT GGCGGATGCG TGTGCGGTGG TGGGGGCGCG
69051 GGCGCGTTTG ATGGGTGGGT TGCCTGAGGG TGGGGCGATG TGTGCGGTGC
69101 AGGCCACGCC CGCCGAGCTG GCCGCGGATG TGGATGGCTC GTCCGTGAGT
69151 GTGGCGGCGG TCAACACACC TGAATCGACG GTGATTTTCA GTCCGTGCGG
69201 TGAGGTGGAT CGGATTGCTG GGGTGTGGCG GGAGCGTGGG CGTAAGACGA
69251 AGGCGCTGAG CGTGAGTCAT GCTTTCCATT CGGCGTTGAT GGAGCCGATG
69301 CTCGGGGAGT TCACGGAAGC GATACGAGGG GTCAAGTTCA GGCAGCCGTC
69351 GATCCCGCTC ATGAGCAATG TCTCCGGAGA GCGGGCCGGC GAGGAGATCA
69401 CATCCCCGGA GTACTGGGCG AGGCATGTAC GCCAGACAGT GCTCTTCCAG
69451 CCCGCGCTCG CCCAAGTGGC CGCTGAGGCA CGCGCGTTCG TCGAACTCGG
69501 CCCC GGCCCC GTACTGACCG CCGCCGCCCA GCACACCCTC GACCACATCA
69551 CCGAGCCGGA AGGCCCCGAG CCGGTCGTCA CCGCGTCCCT CCACCCCGAC
69601 CGGCCGGACG ACGTGGCCTT CGCGCACGCC ATGGCCGACC TCCACGTCGC
69651 CGGTATCAGC GTGGACTGGT CGGCGTACTT CCCTGACGAC CCCGCCCCC
69701 GCACCGTCGA CCTGCCCACC TACGCCTTCC AGGGGCGGCG CTTCTGGCTG
69751 GCGGACATCG CGGCGCCCGA GGCCGTGTCC TCGACGGACG GTGAGGAGGC
69801 CGGGTTCTGG GCCGCCGTCG AAGGTGCGGA CTTCCAGGCG CTCTGCGACA
69851 CCCTGCACCT CAAGGACGAC GAGCACCGCG CGGCTCTGGA GACGGTGTTC
69901 CCCGCGCTGT CCGCGTGGCG GCGCGAACGA CGTGAGCGGT CGATCGTCGA
69951 TGCCTGGCGG TACCGGGTCG ACTGGCGGCG CGTCGAGCTG CCGACACCCG

70001 TTCCGGGCGC CGGTACCGGT CCCGACGCCG ACACGGGCCT CGGGGCGTGG
70051 CTGATCGTGG CTCCCACGCA CGGGTCGGGT ACTTGGCCGC AAGCCTGTGC
70101 CCGGGCGTTG GAGGAGGCGG GCGCGCCGGT ACGTATCGTC GAGGCCGGCC
70151 CGCACGCCGA CCGGGCGGAC ATGGCGGACC TGGTCCAGGC ATGGCGGGCA
70201 AGCTGTGCGG ACGACACCAC CCAGCTCGGA GGAGTGCTCT CCCTGCTGGC
70251 TCTCGCCGAG GCACCGGCCA CCAGTTCCGA CACCACTTCC CACACCAGTA
70301 CCAGTTGCGG TACCGGCTCT CTCGCGTCCC ACGGCCTCAC CGGCACCTTG
70351 ACGCTGCTGC ACGGTCTGCT GGATGCGGGC GTCGAAGCGC CTCTCTGGTG
70401 TGCCACGCGC GCGCCCGTGT CGTGCGGCGA CGCCGATCCG CTCGTCTCCC
70451 CGTCGCAGGC CCCGGTCTGG GGA CTGCGGAC GCGTGGCCGC CCTGGAGCAT
70501 CCGGAGTTGT GGGGCGGCCT GGTCGACCTG CCCGCCGACC CGGAGTCGCT
70551 CGACGCGAGC GCGTTGTATG CGGTTCTGCG CGGAGACGGC GGCGAGGATC
70601 AGGTCGCGCT GCGCCGGGGC GCGGTCCTCG GCCGTCGCCT GGTGCCCCGAC
70651 GCAACCCCGG ACGTGGCCCC CGGCTCGTCC CCGGACGTGT CCGGAGGCGC
70701 AGCCCATGCC GACGCGACCT CCGGGGAGTG GCAGCCGCAT GGTGCCGTCC
70751 TCGTCACCGG AGGCGTCGGC CACCTGGCCG ATCAGGTCGT ACGGTGGCTC
70801 GCCGCGTCCG GCGCCGAACA CGTCGTACTC CTGGACACGG GCCCCGCCAA
70851 CAGCCGTGGT CCCGGCCGGA ACGACGACCT CGCCGCGGAA GCCGCCGAAC
70901 ACGGCACCGA GCTGACGGTC CTGCGGTCCC TGAGCGAGCT GACAGACGTA
70951 TCCGTACGTC CCATACGGAC CGTCATCCAC ACATCGCTGC CCGGCGAGCT
71001 CGCGCCGCTG GCCGAGGTCA CCCCCGACGC GCTCGGCGCG GCCGTGTCCG
71051 CCGCCGCGCG GCTGAGCGAA CTCCCCGGCA TCGGGTCAGT GGAGACCGTG
71101 CTGTTCTTCT CCTCCGTGAC GGCTTCGCTC GGCAGTAGGG AGCACGGCGC
71151 GTACGCCGCC GCCAACGCCT ACCTCGACGC CCTGGCGCAA CGGGCCGGTG
71201 CCGATGCTGC GAGCCCCCGG ACGGTCTCGG TCGGGTGGGG CATCTGGGAT
71251 CTGCCGGACG ACGGTGACGT GGCACGCGGC GCCGCCGGGC TGTCCCGGAG
71301 GCAGGGACTC CCGCCGCTGG AACC GCAGTT GGC GCTCGGC GCCCTGCGCG
71351 CGGCGCTCGA CGGGGGCAAG GGGCACACGC TGGTCGCCGA CATCGAGTGG

71401 GAGCGGTTTCG CGCCGCTGTT CACGCTGGCC AGGCCCCACCC GGCTGCTCGA
71451 CGGGATCCCC GCGGCCCAGC GGGTCCTCGA CGCCTCCTCG GAGAGCGCCG
71501 AGGCCTCGGA GAACGCCTCG GCCCTCCGTC GCGAACTGAC GGCCCTGCCC
71551 GTGCGGGAGC GGACCGGGG ACTTCTCGAC CTGGTCCGCA AACAGGTGGC
71601 CGCCGTCCTG CGCTACGAGC CGGGCCAAGA CGTGGCGCCC GAGAAGGCCT
71651 TCAAGGACCT GGGCTTCGAC TCGCTCGTGG TCGTGGAGCT GCGCAACCGG
71701 CTGCGCGCCG CCACCGGGCT CCGGCTGCCC GCCACCCTGG TCTACGACTA
71751 CCCCACACCC CGCACCTCG CCGCACACCT GCTGGACAGG GTGCTGCCCCG
71801 ACGGCGGCGC GGCAGAGCTC CCCGTGGCCG CCCACCTGGA CGACCTGGAG
71851 GCGGCCCTCA CCGACCTGCC GGCCGACGAC CCCC GGCGCA AGGGCCTGGT
71901 CCGGCGTCTA CAGACGCTGC TGTGGAAGCA GCCCGACGCC ATGGGGGCGG
71951 CGGGCCCCGC CGACGAGGAG GAGCAAGCCG CGCCCGAGGA CCTGTGACCC
72001 GCGAGCGCCG ACGACATGTT CGCCCTGATC GACCGGGAGT GGGGCACGCG
72051 GTGAGCGGGG TGGAGCGGGG TGTGGGGTCG GCGGGCCCTG TGGAACAGGG
72101 TGACGGACTC GCGGGCCTGG TCGAGCGGGC CGAGGCGCTG GCCGCTCTGC
72151 GGGGCGCCTT CGACGGCTCC CCGGGCACCG GCGGCAGCCT CGTCGTGCTC
72201 AGCGGCGCGG TGGGCACCGG CAAGACCGCG CTGCTACGGG CGTGGGCCGA
72251 CCGCATCGGC GCCGATGCCG ACGCCCTGGT CCTGACCGCC ACCGCCTGCC
72301 GCGCCGAGCG CGACCTGCCG CTTGGCGTCC TGGAACAGCT GGTACGCAGC
72351 CCCGGCCTGC CCCC GGCCAG CGCCGAGCGC GCGCTGGCGT GGTGGGACGA
72401 GGAGGCCTCG GCCACCCCCG GAAAGACGGA CGCGAACGGG ACGAGTGCCA
72451 ACGGGACGGA CGCCAACGGG ACGGGCGCGG GACAGACGGG CGCGGGGCAG
72501 GCGGGCGTGG GACAGACGGG CGTGGGCGGA GAGCCCGTCC TGGCCGCCTC
72551 CGCCCTGCGA GGCCTGTGCG AGGTGCTGCG GGACCTGCTC GCCGAGCGGC
72601 CCGTCGTGGT CGCCGTCGAC GACGCGCACC ATGCCGACGC GGCGTCGCTC
72651 CAGTGCCTGC TCTCCGTGGT GCGCCGGCTG CGGTCGGCAC GACTCCATGT
72701 GCTGTTACCC GAGTACGCCC ATCAGAAGGC GCAGAACGCC CTGCTGAGCA
72751 GCGAGTTCCT GCACGAGCCC GCCCTGCGGC GGATCCGCCT GGAACCGCTG

72801 TCGAAGGCGG GCGTGGAGGC CTTGCTCGCC CGGCACCTCG ACGAGCGGAC
72851 GGCACAAGAC CTCACCCCCG TCGTCCACGG CATGAGCGCG GGCCACCCCG
72901 TCCTCGTACG GGCGCTGGCC GAGGACCACC GTGCGGCGGG CGGCGCCGGG
72951 GAGGCGTACG GTCGTGCCGT CCTCAGCTTT CTGTACCGGC ACGAGACTCC
73001 GGTCACCCAA GTCGCCCCGCG CCATCGCTGC GTTGGGCGCG CACGCCGGAC
73051 CCGGTCAGGT CGGGCGGCTG CTCGATGTCG ACGCGGCGTC CGTCGAGCGG
73101 GCCGTGCGGC AGCTGACCGT CGCGGAGGTG CTGCACGAGG GCCGCCTGTG
73151 CCACCCGGCG TTCGCGGCGG CGGTCCCTGGA CGGCATGCCG CCCGAGGAAC
73201 GCCGCGCCCT GCACGGACGG GTCGCCGACC TCCTGCACGA GGAGGGGGCG
73251 CCGGCCACCG AAGTGGCCGC CCACCTCGTC GCCGCCGACC GGTCCGACGC
73301 CCCGTGGGCG GTACCCGTCT TCCAGGAAGC GGCCCAACTC GCCCTGGACG
73351 AGGACCAGGT GGAGACCGGC GTCGACTATC TGC GCGCGGC CCACCAGCGG
73401 TGCCGGGGCG CCGCGCAGCG TGCCGCGGTC GTCGGTGCGC TCGCCGACGC
73451 CGAGTGGCGG CTCGACCCAG CAAAGGTCCT GCGCCACCTG CCCGACCCTG
73501 CAGCCATGGC CCCACAAACG GACCCTGCCG CCCTGGCCCC ACACACGGAC
73551 CCCGCACCCA CAGCCGCACC CACAGCCGCC CCCACCCCCA CCCCATCCC
73601 GACCACCCA CCCCTCCCCA CCCACCTGCT CTGGCACGGG CGGGTCGAGG
73651 AAGGCCTGGA CGCCATCGGC ACGCTACCG GGCCCGGACC CAACCCGGCG
73701 GGTGCGCCGC CGATGAACCC CGCGGACCTG GACACCCCAT GGCTGTGGGG
73751 CGCCTACCTC TATCCCGGGC ACGTCAAGGA GCGCCTGGGA TCCGGCGCCC
73801 TGTCCCCGCA GCGCTCGACC CCGCCGGCGG TCACGCCGGA GCTCCAAGGC
73851 GCGGGCACGC TGATGAACGA CCTGCTGCAC GGCGGCGAAC GCGACGCCAC
73901 CGAGGCCGCC GAGCGCGCCC TCAACCGCTA CCGGCTCGGC CCCCGCACCA
73951 TCGCGGTCCA GACGGCCGCG CTGGCCGCCC TCACCTACCG CGACCGGCCG
74001 CACCGCGCGG CCGCCTGGTG CGACGGCCTC GTCGCCCAGG CCGACGAGCG
74051 CAACAGCCCC ACCTGGCGGG CCCTGTTTAC CGCGTGCGGT GCCCTGCTCC
74101 ACCTGCGGCA GGGCGACCCG GCCGCAGCGG AACAGCGCGC CGAAACCGCC
74151 CTCGCCCTGC TCGGATCGAA GGGCTGGGGC GCCGCGATCG GCCTGCCGCT

74201 GGCAGCCGCC GTACAGGCCA AGGCGGCCCT CGGCGATGTC GACGGGGCGG
74251 CGGCCCTCCT GGAACGGCCC GTGCCCCAGG CGGTCTTCCA GACCCGCACC
74301 GGA CTGCACT ACCTGGCGGC CCGGGGCCGC TATCACCTCG CCACCGGCTG
74351 CCACTACGCC GCACTGTGCG ACTTCTACGC CTGCGGGACC CGCATGAGCA
74401 GCTGGGGAGT GGACCTGCCC GCGCTGGAGC CGTGGCGCCT CGGCGCGGCG
74451 GAAGCGTACC TGGCCCTCGG CGAAGGACTC CTGGCACGCC AACTCGTCGA
74501 CGGCCAGCTG CCGTTGCCCA CGCCTGACGA CGGCCGCACC TGGGGCATGA
74551 CGTTGCGCCT GCGGGCGGCC ACGTCCCCCG CGCCGGCCCCG GGCCGAAGTC
74601 CTCGACGAGG CCGTGGCGGT GCTCCGGGAG AGCGGCGACA CCTTCGAGCT
74651 GGC GCGGGCC GTCGCCGACC AGGCTGTTGC CGTACGCGAA GGGGGCGAGG
74701 CGGAACGCGC CCGGCTGCTG GCCCGCAAGG CGGAGCTGCT GGCCCGGCGC
74751 TGGGGCAGCG CCCCCGCGCC CGCCACCGTC CCCGAACCGC CGGAGCGGCC
74801 AGGACCGGCC ACTCCGGACG CCGAACTGAC CAGTGCGGAG CGGAGGGTGG
74851 CCGAGCTGGC CGCCGAAGGG TTCACCAACC GGGAGATCTC CCGGAAGCTG
74901 TGC GTACGG TCAGCACCGT GGAACAGCAC CTGACCCGGA TCTACCGGAA
74951 GCTCGACGTC AGGCGACTGG ACCTCCAGGC AGCCCTCGGC TGACCTTCAG
75001 GCGGCCCTCG GCTGACCGCA GGCCACGCGC CTACGGTCAG CCTTCCTGAG
75051 TCAGGACCGT ACAGCCGCCG TAGGTGTAGG TGTAGGCGTG GGCGAGATCG
75101 TCGCCGCGTC CAGACCCACC ACGGCCAGCT CCTCCGGAAG GAACGGGGGA
75151 GCGGTCAGCT CCGGGAGGCG TTCGTCGGCG CGCATCGCCA TCAGGAAACG
75201 GTTGGAGCCC AGTTCGGCCT GCGGCGCGTT GAGGCTCATC ACGTCCGTGA
75251 CGATCTCGGA CGCCTTCGGG GAACGGATCG ACGCCGCGGT GATGGCCTCG
75301 GCGAACCGCA GACGCTGCTC GGTGTCCACA CCGATGAGCC GCGGATCCGT
75351 CGCCGAGACA CGGCAGTTGA CGTAGTCGAT GTCCTTGGTC GCGGCGAGGA
75401 TCCACGGGTC GTCCACGGCC GCGCCGATCG CCTTCTGCAG GGCGCGGGTG
75451 CCGGCGCGGG CGGACCCCGT ACCCTCCTGC ACGCTCCGCT CGAACTCGCG
75501 GTCGATCGTG GTGGCGCAGC GCGCGGCCGA GCTCATGCCG TGGCCGTAGA
75551 TCGGGTTGAA AGCGGTCAGC GAGTCGCCGA TGACGAGCAG ACCGTCGGGC

75601 CACTGTTCGA GGCCTCCGG ATAGAGGCGG CGGTTGGCGC CGGAGCGGGA
75651 ACCGAAGACG GGGGTGAGTG GTTCGGCGTC CCGGAGCAGG TCGGCGAGGA
75701 TCGGGTGGTT CAGGTTCTCG GCGAAGGGGA TGAAGTCGTC CTCGTGTGTG
75751 GGCAGTTGCG CGCCCCGCGT GCAGGAGAGC GTCGCGAGCC AGCGGCCGCC
75801 CTCGATGGGG TAGACCACGC CGAAGCGGCC GGGTTCGCGC ACCCGGTCGT
75851 CGGCGGCGAT GTTCACGGCG GGGAAGTGCG TCGTAGCGCC CGGCGGGGCC
75901 TTGAAGAGCC GGGTGGCGTA GCGACGCCC GCGTCCACGA CGTCTTCCTC
75951 CAGTGCCGGC ACGCCGAGGG CGGCGAGCCA CTGCTTGAGG CGGGAGCCGC
76001 GCCCGGTGGC GTCGATCACC AGGTCGGCCT CCAGCTGCTC CTGCCGACCG
76051 CTGTGAGAGT CGCGGACGAC GACACCGGTG ACCCGGCCGC CACTGCCACC
76101 ACCACTTCCC GTCAGCTCGA CGGCCTCGGT GCGCTGCCGG ACGGTGATGT
76151 TGTCGGCTCC CAAGGCCTGC TGACGTACCG TCAAGTCCAG CAGCGGGCGG
76201 CTGGCGACCA GCGCGAACTG GGTGGCGGGG AAGCGGTGCT GCCACCCCTG
76251 ACCGGTCAGC GTCACCAGGT CCTCGGGGAA GCCGAGGCGG CGGGCGCCGG
76301 CCGCGAGGAG GCGGTCGGTG GTGCCGGGCA GCATCTCCTC GATGAGGCGG
76351 GCGCCGTTGG ACCACAGGAG GTGCGCGTGG CGGGCCTGCG GGACCCCTT
76401 GCGGTGCTGG GGCTCCTCGG GCAGCGCGTC ACGTTCCACG ACGGTGACGG
76451 CGTCGACGTG CCGGGCCAGG ACGTGGGCCG CCAGGGTGCC TGCCATGCTG
76501 GCACCCAGGA CGACGGCATG TGCGGGTCGG GTGGTGGTCA CGCGCGTATC
76551 CCTTCGGGGT GGGTGGTGTC GCGGGGCCCG GCCGATCGT CCATGGTCAC
76601 GTCCGTGACG CCCAGAACG CCTGGACCCG GCGGCCGAGC CCGTGCTCGT
76651 CGAGTTCGAC GATGCCGACG ATGCGGAAGG TCATCGGCCG CGGCCGCTGC
76701 ACGGTGACCG TGGTCGGCGT CACCACGAAA CGGTCGTCCA TCGACGTCAT
76751 CGGCGGGTCC GGCACCTCGT GCGTACCGCA GGAGACGGCC AGTTCGAGAT
76801 GGCGGCGGAG ATCGTCCTTG CCCACCATCG GGGGCCGCC CACGGGGTCC
76851 TCGAAGACGA TGTCGTCCGT GAACAGGTCG AGGACGCCTT CGATGTCACC
76901 GCGGTTGATG CGCTCGGCGT AGTCGACGGC CATCTGCTTG CGCGCGGCCT
76951 CGTCGGGCAT GGCACCTCCA GGAAGGGTGG GCAGACCTTG TGAAAGTCAT

77001 CGAGGGCCGT TCGGTTTCAGC CGAGGACCGT GAGATCGGAT GTGCCCCAGT
77051 ACGACTTCAG ATGCCGGATG AGGCCGGACG CGTCCATGCG GATCACGAGC
77101 ATCGCCGTGC GGTGTATGCG GGCCGTCCCC GGGGCGTCGG GGGCCTTGAG
77151 CCAGCCCCGC TCCGCGTAGA GCGGGCCCAC GGGCAGGTAG TCCATGACGG
77201 AGGAAATCTG GATCAGCGCG TCGGTGGCGT CCTGCCCCGC GACGGGCTCG
77251 GCCGCCTCCT CGCGCAGGTG CGCGGCGAGC AGCGGTTTCGT AGTGGGCGCG
77301 CAGCGCGTCG TGCCCGGTGA CGGGCGGGAG GCCGACCGGG TCCTCGAGGA
77351 CCGCGTCGGG CGCGTACAGA TCGATGATCG CGTCCAGGTC CCCGGCGTTG
77401 ATCCGCCGGC TGTGCTCCAG GGCCCCCTTC TTGCGGGCGA ACTCGTTCAT
77451 CGCTGCCCCCT CCACTGCCTG ACCGTGTCCG TTGCCGTTGC CGTTGCCGTT
77501 GCCGTTGCCG TGTCGTTGCG CCTGCCCCGT GGGCTGTCCG TTGCCCTGTC
77551 CGCTCGCGCC GTCCCTGCCG AGGTCCCCGT CGATGAACGC GAAGATCTCG
77601 TCCGCCGACG CGTCCTGGAT ACGTGTACGA GTGGCCACCG CGACCTCGCC
77651 GGCCGTGTCC TGCGGCGCGT CGAGCCTGGC CAGCGTCGCG CGCAGCCGCC
77701 CCGCCAGTTC GGCCCGCGCC GAGCCGTCCT TCGAGGAGAC CGAGAGCAGC
77751 GAGTCCTCGA TGCGCTCGAA CTCCGCCAGG ACGTCGGCGA GCGGATCCGC
77801 CGCGCGCGGG GCCAGCTCCT GCCGCAGCTG CGCGGCGAGC TCCGCCGGGT
77851 TGGGATGGTC GAAGACGAAC GTGGCGGGCA GCTTCAGCCC CGTCGCGGCC
77901 GAGAGCCGGT TGCGCAGCTC CACCGCGGTC AGGGAGTCGA AGCCGAGTTC
77951 CCGCAGCCCC TCGTTCGCGT TGACGGGCGT GGCCGCGTCG TAGCCGAGGA
78001 CGGCCGCGAT ATGGGTGCAC ACCAGGTCGA GCAGCGCCTC CTCCCCGCTCG
78051 GGGTCGGACA TCGCGCCGAG CGACTTGAGC AGCGCGGCCG CCCCCGCCGA
78101 CACGGCACCG CCGCCGCTCT TGCTCCCCC GCGCACCAGG TCGCGCAGCA
78151 GCGCCGGTGC GGGGTGGCTC TGGGCCTGCC GGCGCATCCG GGCCAGGTCC
78201 AGACGGACCG GCGCGTACAG GGGCAGTCCG CCGGCCACG CCGCGTCGAG
78251 GAGGGCGAGT CCCTCGTCGG CGCCGAGCCC GACCACGCCG GCGCGGGCAT
78301 GGCGCGCCCC GTCGGCGTCG GTGAGCCGTC CCGACATGCC GCTCGCCAGC
78351 TCCCAGTAGC CCCACGCCAG GGAGGTCGCC GCCGCACCGC CGTCGTGCCG

78401 GTGCCGGGCC AGCGCGTCCA AGAAGGCGTT GGC GGCCGTG TAGCTGCCCT
78451 GGCCGGGGCC GCCGAGCAGC CCGGCGACCG AGGAGTACAG GACGAACGCG
78501 GACAGGTCCG CGTCCCGCGT CAGCTCGTGC AGGTGCCACG CGGCGTCCGC
78551 CTTACGCGC ATCACCTCCT CGACCTGCTC GGCCGTGAGG TTCTGCACCA
78601 CGGCGTCGTT CACGGTGCCC GCGCAGTGA AGACGGCGGT CAGCGGGTGG
78651 TCCGAGGGCA CCGCCGCGAG GAGGGCGGCG GCTTCGTCCC GGTCGCCCCG
78701 GTCGCACGCG GCGAAGGTGA CTCGCGCGCC GAGCGCGGAG AGGTCGGCGG
78751 CCAGTTCGAG TGCGCCCGGC GCGTCGGCTC CCCGCCTGCT GGACAGCAAC
78801 AGGTGCCTGG CTCCGTACCG TTCCACCAGG TGACGGGCCG TCAGCGAGCC
78851 GAGTGCTCCG GTGCCGCCGG TGACCAGCAC GGTGCCCTCG GGGTCGAAGG
78901 CGGGAGGCAG CGAGAACACG GTCGTGCCCC CCGAGGGCGG GGCCGCCATC
78951 GCGGCGGGCG CCTGCCGGAT GTCCCACACG GTGATGTCGA GCGGCGTCAG
79001 AGCACGGCTG TCACCCCGTT CCGCGGGCAG CCCGGGCTCC GCCGACTCCG
79051 TGATCTCGGC AAGCTCGGTC AGCTCGGTCA GCTCCGCGAG GATTTCCCGT
79101 ACGCGCCCGG GCTCGGGCGG CACGACAGCC TGTCCCTCGT CCGGACGACC
79151 GCCCGCACGG TGGACCACCA GGGCCCCCTC GTGGCGGAGG GTGACGTCGG
79201 CCGCGCGCTC GGCGCCCGAA TCATCCGCCG TCGACGCACC GTCCACGGCC
79251 TCGACCCGCC ACCGCCCGGC AAGAGCAAGA CGCAGCACGG CACGGCCGAC
79301 GGAACCGGTT TCCTCCCCGA CGAGCAGAGT CTCGCCGCCG CGCGGCGCCA
79351 CGACATCCGC CAGCACGTGA TACGCGGACA CATAGGCCCC CAAGGACCCG
79401 GCCGCCTGCG CCCAACTCCA GCCCGCCGGA ACCGGCATAA GCAGCGCGGC
79451 ATCGGTGACG GCCACCGGGC CCACCGCGTC GAACAACCCC ATCACCCGGT
79501 CGCCACGGC CACCGAACCG ACCTCGCCGC CGACTTCCGT CACCACACCG
79551 GCACCCTCGA CCTGGCCCCG CGTGAGCGGC CCCGGCGCCG CGGCCCGCAC
79601 CGCCACCCGC ACCTCGTGCG GCTCCAGCGC CCGTCCGGCC TCGGGAGCGT
79651 CGACAAGGGA CAACTGCTGT CCGCCGCCCC CCTCTTGGA CCGAGCCAGC
79701 CGCCACGTGA GCGATCCGAC CGGCGGCACC AGCCGCACCG ACGCGTCGTC
79751 GCGCACGAGC CGTGGCACGT AGGCGCGCCC GTCACGCAGC GCCAATTCCG

79801 GTTCGCCGGA GGCCAGTACG CCGGTCAGCG TGGCCGGAGA AGACTCCAGT
79851 CCGTCCACGT CGAGCAGCGT GAGGCGACCG GGATTCTCGG CCTGCGCGCT
79901 GCGCACCAGA CCCACAGCG ACGCGCCCGC CAGATCACCG GCGGTCTCAC
79951 CCGGCCGCGC GGCGACCGCG CCTCGGGTGA CGACGACGAG ACGGGTCGCC
80001 GCGAACGCCG GGTCGTCCAC CCACTCCTTG AGCAGCGACA GAAGGGACAC
80051 GGTGGCCAGC CGCGCGTACC CGGCCGGGTC GCCGCCCCTG CCATCGGCAT
80101 CCGCAACGGC CCCGGCACCT GCGCCGGGCG CGGCGCACAC GGCGAGCACG
80151 ACATCGGGCG CTTCGCCCCC AGCCGCCACT CCGTCCCGGA GCGCACCGAA
80201 CGTGTCCAC ACGGGGCCGG CGGCCAGCGC ATCGGACAAG GCGTCGGCCA
80251 GCGCACCGGC CGACGTACCG CCCATCGGGC CACTCTCGAC CGGCGCGAGG
80301 ACCGCGGCAC GCGGGGCGCC GCCGCCCGTC TCCTCGGCC GCGCGGCGAC
80351 CTCCATCCAC ACGAGCCGGA ACAGCGCGTC ACGGTCCGCC GCACGGGCGC
80401 CCGCGATCTG GTGGGCGGCC ACCGGCCGTA CCGTGAGCGA CTCCAGCGTG
80451 AGAACCGGCT CCCC GCCTCC GCCCCCGTCC ACGGCCGTGA GGGCCAGCTG
80501 GTCGGGCGCG GTGCGTGCGA TACGTACCCG CAACTTCTCA GCGCCCGGCG
80551 CGTGCACCCG CAACCCGCTC CAGGAGAACG GCAGCAGCAC TTGGTCGGTG
80601 TCGGCGGACG ACGTGACCGC GTCCAGGATC AGCGCGTGCA GCGTGGCGTC
80651 GAGCAACACC GGGTGCACCT GGTAGCGGTC GGCCCTGCCG CTCTCCGCCT
80701 CGGGCAGCGC CACCTCGGCG AAAAGGTCGT CCCCAGCCG CCACGCGCTC
80751 ACCAGTCCCT GTGAGCCGGG CCCGAAGTCA TAGCCGTACG AAGCGAGTTC
80801 CCCGTACGGA TCCTGCTCGC CGACCGGTGT GGCGCCCGGG GGCGGCCACG
80851 TCCCGCCGAA CGAGGCGTCC CCGGCGTCGG GCCCCGGGG AGCGACCACG
80901 CCCGCGGCAT GCCGGGTCCA CACGGCCTCC TCGCCCTCAC CCGTGGGCCG
80951 CGAATGGACG GTCACGGGAC GCCGCCCGTC CTCGGCCACG GAACCGACCA
81001 CCACCTGCAC GTCGACCGCG CCCGCACCCT CGTCCCCGAA GGCGAGCGGA
81051 GTGTGCAGCG TCAGCTCCGC CAACTCCGCG CAGCCGGCCC GCACCGCGGC
81101 CTGCAGCGCG AGCTCCACGA ACGCCGAACC GGGCAGCAGC ACCGTGTCCA
81151 TGACCCGGTG CTCGGCCAGC CACGCCTGGT CCCGCGGAGA GATCCGGCCG

81201 GTCAGCAGGT GACTGCCGCC GTCCGCGAGT TCCACGGCGG CTCCGAGCAG
81251 CGGATGCCCC GCGGACGCGA GCCCGAGCCC CGCCGGGTCC CCGGCGAGCC
81301 CCCTGCGCCC CTCCAGCCAG AACCGCTCCC GCTGGAAGGC GTACGTCGGC
81351 AGATCCACCA CCCGAGGCAG CGGCACGGCC GGGAACCAGC CCGTCCAGTC
81401 GACCTCCGCC CCCGCGCCGA AGGCCTGGGC GGCCGCGCGG GTGAGCTGCG
81451 CGGCGTCGCC GTGGTCGCGG CGCAGGGTGG GCACGACGGT GGCGGGCATG
81501 TCGGCCCCGCT CGATGGTCTC CTCCATGCCG AGGTTGAGGA CGGGGTGGGG
81551 GCTGGCCTCG ATGAACAGGC GGTAGCCGTC GGCCAGCAGC GCTTCGATGG
81601 TGTCGGCGAA GCGGACGGGC TGGCGGAGGT TGGTGACCCA GTAATCCGTG
81651 TCGAGGGTGG TGGTGTCTG GAGGCGTTTCG GCGGTGACCG TGGAGTAGAA
81701 GGCGACGTCC GTGGTCGTGG GCCGGATGTC GGCCAGGCGC TCGGTGAGGA
81751 GGTCTGTGAG CTGGTCGATC TGGGGGCCGT GGGAGGCGTA TCCGACGTCTG
81801 ATGACGCGGG CGCGCAGGCC TCGCGCCTCC GCATCCGCGA CCACGGCTGC
81851 CACATGCTCC GGCGGCCCTG AAATGACCGT AGAGGAGGGC CCGTTGACGG
81901 CAGCGACACA CACGCCGGGC CGGTGCGCGA TGAGCTCAGC AACCTGCTCC
81951 GAGCCGGCCC CCAACGACGC CATGTGCGCC TGCCCCATGA GCTGACGGAG
82001 CGCGTCACTG CGTACGGCCA CGATCCGCGC CGCATCCTCC AGTGACAGTG
82051 CCCCCGCCAC ACACGCGGCG GCCATCTCGC CCTGCGAGTG CCCGATGACG
82101 GCAGCCGGGG TGATGCCGTA ATCGGCCAC ACCGAAGCCA GCGAGACCAT
82151 CACCGCCCAC AACACGGGCT GCACGACCTC GACCCGGGAC AGCTCACTCC
82201 CGTCCCCGCG CAACACCGCA CTCAGCGACC AGTCCACATG CGCCGACAGG
82251 GCCCCTCAC ACTCCGCGAT CCGCGCCGCG AAGACGGGGG ACTCGTCAAG
82301 GAGCTGGGCA CCCATGCCCA CCCACTGCGA CCCCTGCCCC GGAAACACCA
82351 ACACCGGACC CGCGCCGGAG GCGCCCTGTA CGGCGCCCTC GACGACGTCC
82401 GGTGACGGCT CGCCCGCCGC CAGGGACCGT AGCCCGGCGA GGAGAGTCTG
82451 GCGGTCCTTG CCCACGACGA CGGCTCGGTT CTCGAACACC GACCGGGTCT
82501 TGACCAGGGA CCAGCCCACG TCCAGCGGCG ACGCGAGCCG CGGGTCGGCG
82551 GTGGCGCGGT CGGCCAGCAG GCGGGCCTGG GCCCGCAGCG CCTCCTCGCC

82601 GCGCGCCGAC ACCACCCAGG GCACCACTCC GGCCGGCGCC GCGGCGTCCT
82651 CCGCCGGAGC GGTACAGGGC TCCGGCGCGT CCGGGGCCTG TTCCAGGATG
82701 AGGTGCGCGT TGGTGCCGGA GATGCCGAAG GCGGACACCC CGGCGCGGGC
82751 CGGGCGTTTCG CCGCGCGGCC AGGAGACCGG TTCGGACAGC AGGCGGACGC
82801 CACTGCCGTC CCAGTCCACG TCGGGCGTGG GCGCGTCGAT GTGCAGGGAG
82851 GCGGGCAGCT GTTCGTTGCG CAGCGCCATG ACCATCTTGA TCACACCGGC
82901 GACACCGGCC GACGCCTGCG CGTGCCCGAT GTTCGACTTG ATCGAGCCGA
82951 GCCACAGCGG CCGGTCCGCG GGCCGCTCCT TGCCGTAGGT GGCGACGAGC
83001 GCGCTGGCTT CGATGGGGTC GCCCAGCATG GTGCCGGTGC CGTGCGCCTC
83051 CACCGCGTCG ACGTCCTCGG CGGAGAGCCG CGCGTTGGCG AGTGCCTGCC
83101 GGATCACCCG CTGCTGCGCC TGCCCGTTGG GTGCCGTGAG CCCGTTGCTC
83151 GTGCCGTCTT GGTGATGGC CGAACCCCGG ATCACCGCCA GGACGTTGTG
83201 GCCGTTGCGC CGGGCCTCCG AGAGCCGTTT GAGTACGACC AGGCCGACTC
83251 CCTCGGCCCA GCCGGTGCCG TCGGCGGCGG CCGCGAACGG CTTGCACCGG
83301 CCGTCCTTGG CGAGCCCGCG CTGCAGCGAG AACTCGACGA ACGAGCCCGG
83351 CGTGCCATC ACCGTCGCGC CGCCGCGAG AGCGAGCGAG CACTCGCCCT
83401 GGCGCAGCGC GTGCGCCGCC TGGTGGATCG CCACCAGGGA CGAAGAGCAG
83451 CCGGTGTCTGA TCGTCATGGC GGGGCCTTCT AGGCCGAGTA CGTACGACAC
83501 CCTGCCGGAG GCGACACAGC CGAGGTTGCC GGTGCCGATG TAGCCCTCGA
83551 CCTCGGTGGG CTGTTACCG ACAGAGCGCA GGTAGTCGAA GATGGTCAGG
83601 CCCGTGAACA CCCCGGCGTC GCTGCCCTTG AGGGTCTCCC GGTGAGGCC
83651 CGCGCGTTTC ATCGCCTCCC ACGCGGTCTC CAGGAGCAGC CGCTGCTGCG
83701 GGTCCATCGC GACGGCCTCG CGGGGGCTGA TGCCGAAGAA TCCGGCGTCG
83751 AAGTCGCCCC CGTCGTAGAG GAACCCGCCT TCGCGCACAT AGCTGGTGCC
83801 GCGGCTCTCC GGGTCCGGGT CGTACAGCGT CTCCAGGTCC CAGCCCCGGT
83851 CGTCGGGGAA GGCCCCATG GCGTCCTTGC CGGCCGCGAC CAGATCCCAC
83901 AGTCCTTCGG CGGAGCGGAC GTCGCCCAGA TAGCGGCAGG CCATGCCGAC
83951 GATCGCGATC GGCTCGTCGT CGGCGGCGCC CCTGGAGGCC CCGGCCGCCC

84001 GCACCGGGTC GCGGAGGCC GCCGCGTCAC CGGACAGCTC GGCCCGCAGG
84051 ACGTCGGTGA GCGCGTCGGG GGTGGGGTGG TCGAAGACGA CCGTGGTCGG
84101 CAGTGTCAGG CCGGTGCTCT TGTTTCAGCCT GTTGCGCAGC TCCACCGCGG
84151 TCAGCGAGTC GAAGCCCAGC TCCTGGAACG GCTTGGTGGC GGGCACCGCG
84201 TCGACGTCCG AGTGCCCCAG CGTGCCCGCC GCCTGGGAGC GCACGTGCTG
84251 CAGCAGCAAC TGCCGCTGCT GCGCCGGCTT CGCCTCCGTC AGCTCCTGCT
84301 GGAGCGACGA TGCCTCCGTG GCGTCTTCCT GCTGTGCCGC GGGTGCGCTG
84351 GCCCGCCGGT TCTCGGGCAG ATCGGCGAGG AGCGGGCTGG GCCGCTGCGC
84401 GGTGAACGTC GACGTGAACT GCGCCCAGTC GAAGTTCGCC ACGGTCAGCG
84451 TCGTCTCACC CGCGTCCAGG GCCTGCTGCA GCGCCTTGAC GCACAGCTCC
84501 GGGCTGAGCG GGTGCAGGCC GAAGCGGCTG AAGAACGTCA ACGCGGCCTG
84551 GTCCGCCGCC ATGCCCGCCT CGGCCAGGG CCCCCAGGCG ATGGAGGTGG
84601 CGGGCAGGCC CTCGGCGCGG CGGTGCTCGG CGAGGGCGTC GAGGAAGTGG
84651 TTGGCCGCAC CATAGGCGCC CTGCTGGCCA CTGCCCCACA CGCCTGCGCC
84701 CGACGAGAAC ATCACGAACG CCGAGAGCGG CAACTCCCGG GTCAGTTCAT
84751 GCAGATGGTG AGCGGCGAGC GCCTTCGGAC GCAGCACCTC GTCCAGCTCG
84801 GCACCCGACA CGTCGCCGAG ACCGATGTAG TTCGGCACGC CGGCCGCGTG
84851 GATGACGGCG GTCAGCGGGT GCTCGGCGGG GACATCGTCG ATGAGGCGTC
84901 GCACCTGCTC GCGGTCGCCG ACGTCGCAGG CGGTGACGGT GACGGCGGCC
84951 CCCAACTCCG TCAGTTCGCG GCGAGTTCC TGTGCTCCCG GGGCGTCGGG
85001 GCCGCGGCGG CTGGTCAGGA GGAGGTGCGG GGCGCCCGCA CGGGCGAGCC
85051 ACCGCGCGAG GACGGCGCCG ATGCCGCCGG TCCCGCCGGT GATGAGAGTG
85101 GTGCCGTCGG GCCGCCAACC AAGCCCCTG CCGACCGTGT TGGCGGGCGC
85151 GTGTGCAAGG CGACGGGCAT GGACGCCGGA CGGCCGGATG GAGATCTGGT
85201 CCTCGTCCTG CGGAACCAGC GCGGCGGCCA GCCGGGCCAG CGTCTGATGG
85251 TCGATACGAG CGGGCAGATC GACCAGCCCG CCCCACAGCC GCGGATACTC
85301 CAGCGCAGCG ACGCGCCCCA GCCCCACAC CTGAGCCTGC ACCGGGTGGG
85351 TGAGGGCGTC GCCGGCGCTC GTGGAAACAG CCCCTGCGT GAGAGTGCCT

85401 ACGGCGATGT CGGCGCCGTT GTCCGCGAGG GCCTGGACGA GAGCGGTCGT
85451 CGCGGCGAGT CCGGCGGGCA CGGCCGAGTG CTCGGGATGC GGCTCCTCGT
85501 CCAGGGCCAG CAGATTGACG ACTCCGGCAA ACGCGGCCCC GTCCATCAGG
85551 ACACGCAGCT CCTGCGCCAA CTCCGTACGC TCCATGGCAC GTGCGTCGAC
85601 CACGTGGCGT CGCACCTCGC CACCATGGGC GGTCAGCGTC TGC GCGGTGCG
85651 CGAGGACGGC CGGGTGGTCG GCGTGC GCGG CGGGCACGAG CAGCAGCCAG
85701 GCCCCGCTGA GCTCCGGCGC CGGCACGTCG GGCAGATGCT TCCAAGTGAC
85751 CTGATAACGC CAGGAGTCGA CGGTGGACTG CTCGCGGTGC CGACGCCGCC
85801 AGGCCGAGAG GACGGGCAGC GCGGACTCCA GCGCTCCGAC GCTCTCCGCC
85851 TGCCCCCTGA TCTCCAGACT GCCGGCGAGG GCGTCGATGT CCAGGTCCTC
85901 GATCGCCTGC CACACCCGGG CCTCGACCGG ATCGTGCCCA CCACCCACGG
85951 CTGCGACCGC CGCGGGCGGC TCCACCCAGT AGTGCTTGTG CTGGAAGGCG
86001 TAGGTGGGGA GGTCGACGGT ACGGGGGGTG GGGTCGGCCG GGAACCAGCG
86051 CCGCCAGTCG ACGGGGGCGC CGGCGGTGAA GGCGTGGGCG GCCGCGCGGG
86101 TGAGCTGGGT GGTGTCACCG TGGTCGCGAC GCAGGGTGGG GATGGTGACG
86151 GCCGTCCCCG CAGCACCGGC CTGCTGCTCG ATGGTCTCCT GGATGCCGAG
86201 GTTGAGGACG GGGTGGGGGC TGGCCTCGAT GAACAGGCGG TAGCCGTCGG
86251 CCAGCAGCGC TTCGATGGTG TCGGCGAAGC GGACGGGCTG GCGGAGGTTG
86301 GTGACCCAGT AGGCGGTGTC TAGGGCGGTG GTGTCGTGCA GCGCTCTGC
86351 GGTGACCGTC GAGTAGAACG CCACGTCGGT GGTGGTCGGC TGGATGTCGG
86401 CGAGCCGGTC GGTGAGGAGG TCGTGGAGCT GGTCGATCTG GGGACCGTGG
86451 GAGGCGTACC TGACGTCGAT GACGCGGGCC CTGAGTCCCT GCGCCTCCGC
86501 ATCGGCGACG ACGGCTGCCA CATGCTCCGG CGGGCCCGAA ATCACGGTCG
86551 ACGACGGTCC GTTGACGGCC GCGACGACTA CGCCCGGCCG GTCGCCGATC
86601 AGCTCTGCGG CCTGCTCGGC ACCGGTGCTG AGCGAGGCCA TGTCGCCGTG
86651 CCCTTG CAGC TGACGAAGCG CGTCGCTGCG TACGGCTACG ATCCGTGCCG
86701 CATCCTCCAG TGACAGTGCC CCCGCCACAC ACGCGGCAGC CATCTCGCCC
86751 TGCGAGTGCC CGATGACGGC AGCCGGGGTG ATGCCGTAAT CGGCCCACAC

86801 CGCAGCCAGC GAGACCATCA CCGCCCACAG CACGGGCTGC ACGACCTCGA
86851 CCCGGGACAG CTCGCTCCCG TCCCCGCGCA AGACATCACT CAGCGACCAG
86901 TCCACATGCG CCGACAGCGC CTGCTCACAC TCCGCGATCC GCGCCGCGAA
86951 GACGGGCGAC TCGTCAAGGA GCTGGGCGCC CATGCCCACC CACTGCGACC
87001 CCTGCCCCGG AAACACCAAC ACCGGCCCAG GACCCACATC ACCGGCCACC
87051 CCGGCCACCA CATCCGCCGA CGCCTCACCC GCGGCCAATG CCTCCAGGCT
87101 GGCACCAGCC TGAGCCAAGT CCCGCCCCAC GACCACGGCC CGCTGATCGA
87151 ACAACGCGCG TGTCGTGGCC AGCGACCAGC CCACCTCGGA GACCGACGCA
87201 TCCGCCAGCC CGGCCGCGAA CTCGCCCAGC CGCCGCGCCT GTTCACGCAA
87251 CGCGTCCGGC GTCCGCCCCG ACACCACCCA CGGCACGACC CCACCCGGCT
87301 CAGCCGCCAC GGGGCCCCGG GCGTCCTCTT CCGGCGGCGC CTCCTCCAGA
87351 ATCAGGTGCG CGTTCGTCCC GGAGATCCCG AACGCCGAGA TGCCTGCCCCG
87401 CCGCGTGCGC TCCGCCGGCC AGTCCACGGG CTCGGAGAGC AGTCGTACGC
87451 TGCCCTGTTC CCACTGGACG TGCGGTGACG GGGCGTCGAT GTGCAGGGAG
87501 GTCGGGAGGA GACCGTTGCG CATCGCCATG ACCATCTTGA TGACGCCCCG
87551 GACACCGGCG GCGGCCTGCG TGTGGCCGAT GTTGGATTTC ACCGAGCCGA
87601 GCCAGAGCGG ACGGTCCTCC GGGCGCCCCT GGCCGTACGT GGCGATCAGG
87651 GCCTGCGCCT CGATGGGGTC GCCGAGCGTG GTGCCGGTGC CGTGCGCCTC
87701 TACGGCGTCG ATGTCTCTCG CGGAGAGGCG GGCGTTGGCG AGGGCGGCGC
87751 GGATGACGCG TTCCTGGGAG GGGCCGTTGG GGGCGGCGAG CCCGTTGCTC
87801 GTACCGTCCT GGTGTTGGTC CGAACCCCGT ATCACCGBAA GGACCTTGTC
87851 GCCGCGGCGC CGCGCTTCGG AGAGCAGCTC CAGCGCCACC ACCCCGGCGC
87901 CCTCGCCCCA GCCGGTGCCG TCGGCGGCGG CCGCGAACGG CTTGCACCGC
87951 CCGTCGGGCG CGAGCCCCCG CTGCCGGGAG AACTCGGTGA ACGAACCCGG
88001 CGTCGCCATC ACCGTCGAAC CGCCCGCCAG CGCGAGCGAG CACTCGCCCT
88051 GCCGCAGCGC CTGACTTGCC AGATGGATCG CCACCAGGGA CGACGAGCAC
88101 GCCGTGTCTGA CGGTGACCGC GGGACCTTCG AGCCCCACCG TGTAGGAGAT
88151 CCGGCCCCGAC ACCCACTGCG CGAGGTTGCC GGTGCCGATG TACCCCTCGA

88201 CGTCGCTGGC CGTCTGGCTG ATCAGCGTCA GGTAGTCGTG GCGCTCACT
88251 CCGGTGAAGA CGCCGGTGTC GCTGCCCTTC AGCGCGTGCG GGTTCATGCC
88301 CGCGTGCTCG ATCGCCTCCC ACGCGGTCTC CAGGAGCAGC CGCTGCTGCG
88351 GATCCATCGC CGTGGCCTCG CGCGGGCTGA TGCCGAAGAA CTCGGCGTCG
88401 AAATGGCCGG CGTCGTACAG GAAGGCGCCG TCCCGCACAT AGCTGGTGGC
88451 CGGATGCTCC GGATCCGGGT GATACAGCGA CTCCAGGTCC CAGCCCCGGT
88501 CGTCGGGGAA CCCC CGCACC GCGTCACCCC CGTCGCGCAC GAGTTCCAG
88551 AGGTCCTCCG CCGACCGGGC GCCGCCCGGG TAGCGGCAGG CCATGCCGAC
88601 GATGGCGACC GGCTCGGTCTG ATTCCTTGTC GTGGAGCCGT TGCCGGGCCT
88651 GGCGCAGCTC CGCGGTGACC CACTTGAGGT GATCGAGAAG CTTCTCCTCG
88701 TTCGACATCT GACCCAGGCT CCTTGGCGCT ACGTGGTGAT CGGGGCGTAT
88751 GAGGTGGGG GAGGGCAAGG GGGCCGGTGT GGCCGGGGCT CATCGCGCTC
88801 AGGACTGATC GCTGCTCAGG ACTTCCCGAA CTCACTGGAG ATGAGGTCGA
88851 AGATGTCGTC CGCGCTCGCC GCCTCCAGAT CGGCATGGGC CGAATCAGTG
88901 CCTTCCGGCC CGTCCTGCGC CGGACTCCAC TTCGACACAA GGACCTGCAG
88951 CCGGCCCACG ATGCGGCGCC GGGCCGCCTC GTCCACCTCG GCCGCTCCGA
89001 ACGCCGTGTC CCACTTGTCG AGCGCCGCGA GCACGTCGCC CTCACCTGCG
89051 ACCTCGGCGC CGTCGCCGAG CTGTCCGCGC AAGTGCGTGG CGAGGGCCTC
89101 GGGCGTGGA TGGTCGAAGA TCACGGTGGC GGGGAGCGAG AGTCCGGTGC
89151 TGGTGTTGAG CTGGTTGCGC AGCTGGACCG CGGTGAGCGA GTCGAAGCCC
89201 AGCTCCTGGA ACGGCTTCGC GCGGGAATG TCCTCCACCG TCGGCGCGAG
89251 CGTCGCGGCC GCGTATGTCC GGACCTGCTG GACCAGGAAG CCGAGCCGCT
89301 GTGATGCGGG CGTCTTCGCC AGCTCCTCGC GGAACGCGCT CGTCTCGGCG
89351 GCGGTCCCCG TCTGCTCGGC CTCCCGCTGG TTCTCCGGAA GGTCGTCGAG
89401 GAACGGAATG GGCCGCTGCG CGGTGAACGT CGGCGTGAAC TTCGCCCAGT
89451 CGAAGTTCGC CACGGTCAGC GTGGCGTCGC CCGCGTCGAC CGCCTGGTGC
89501 AGCGCCTTGA CGCACAGATC CGGAGCGATC GGGAGCAGAC CGAAGCGCTT
89551 GAAGTACGTC AGTGAATCCG GGTCGGCGGA CATGCCCCGCC TCGGCCCAGG

89601 GCCCCAGGC GATGGAGGTG GCGGGCAGGC CCTGGGCGCG GCGGTGCTCG
89651 GCGAGGGCGT CGAGGAAGTG GTTGGCCGCA CCATAGGCGC CCTGCTGGCC
89701 ACTGCCCCAC ACGCCGGCGC CCGACGAGAA CATCACGAAC GCCGAGAGGT
89751 CCAGGTCGCG CGTCAACTCG TGCAGGTTCC AGGCCGCGTC GGACTTCGAC
89801 CCCAGCACCT CGCCGAGGCG CGCGGTCTGC AGATCACCGA TCGCGGTCAG
89851 ATCGGTCATG CCCGCCGCGT GGATGACGGC TGTGAGGGGA TGCTCGGCCG
89901 GCATGTCGTC GATGAGGCCG CTCAGTTGGC GGGGATCGCT GACGTCGCAG
89951 GCGGTGATGG TGACGGCGGT GCCGAGCCCG TCGAGCTCGG CGGCGAGTTC
90001 CCGGGCGCCG GGCGCGTCGG GCCCGCGACG GCTGGTGAGG TGAAGACGGG
90051 GGGCGCCCTG CCGGGCCAGC CAACGGGCGA GGACGGCACC GATGCCGCCG
90101 GTCCCGCCGG TGATCAGGTG GGTGCCCCGA GGCCGCCAGG TGGCCTCGCT
90151 GTGCACGGGA TTCTGAATGC TTCCGACGGC GTGCGTGAGG CGCCGGTGGT
90201 GGATTCCGGT GGGGCGGACG GCGGTCTGGT CCTCGTCGTC CTGGGGGAGG
90251 AGAGCGGCGG CGAGGCGGGG GAGGGTGTGG CGGTCGATAC GAGCGGGGAG
90301 GTCGACGAGT CCGGCCCAGA GGCGCGGGTG TTCGAGGGCT GCGACGCGGC
90351 CGAGCCCCCA GACGTGAGCC TGGAGGGGGT GGGTGAGTGG GTCGGTGGCG
90401 CCCGTGGACA CGGCACCCTG CGTGACGGTG TGCAGGGGTG CGGTCGTGCC
90451 GTTGTGCGCC AGGGCCTGGA GGAGAGCGGT CGTCGCGGCG AGCCCCGCGG
90501 GCACGGCGGG GTGCTCGGGG TGCGGCTCCT CGTCCAGCGC CAGCAGATTG
90551 ACGATTCCGG CAAGACCGGC CGTGTCCACC GCGGCCAGCT CCTGACGTCC
90601 CGCCCGGCCG GTCTCGACCG GATGCAGCCG GACGGCGGCC GCCCCGTGCT
90651 CGCTCAACGC CTCGGCGGTG GCTCGTACGG CGGGGTGCTC CGCCTTGTCG
90701 GCAGGGACGA ACAGCAGCCA GTCGCCGCCG AGTTCCGGTG CGGGCCCCTG
90751 GGACCGCTGT TTCCACGTGA CGCGGTACCG CCAGGAGTCG ATGGTCGCCT
90801 GGTCTTGGTG CCGACGCCG CAGCCCTTGA GCACCGGCAA CGCGGGCTCC
90851 AGCGCCCGGA CCGCCTCCTC GCTGCCCTCC TCCGACCCCA GCGTCTCGGC
90901 CAGCAGACCG AGATCGAGCT CCTCGACGGC GTGCCACAGC TGGGCCTCGG
90951 CCGCACTCTG CTCACCGCTG ACGGCGCCCG AGGCGGACGC GGAACGTTCC

91001 AGCCAGTAGT GCTGGTGTG GAAGGCGTAG GTGGGGAGGT CGACGGTGCG
91051 GGGGGTGGGG TCGGCCGGGA ACCAGCGCCG CCAGTCGACG GGGGCGCCGG
91101 CGGTGAAGGC GTGGGCGGCG GCACGGGTGA GCTGGGTGGT GTCGCCGTGG
91151 TCGCGGCGGA GGGTGGGGAC GACGGTGGCG GGGATGTCCG CCTGCTCGAT
91201 GGTCTCCTCC ATGCCCAGGC CCAGCACGGG GTGGGCGCTG GCCTCGATGA
91251 ACAGGCGGTA GCCGTCCGCG AGAAGGGCTT CGATGGTGTC GGCGAACCGG
91301 ACCGGCTGGC GGAGGTTGGT CACCCAGTAA TCCGTATCCA GGGCTGTGGT
91351 GTCCGTCAGA CGCTCGGCGG TGACCGTCGA ATAGAAGGCC ACGTCCGTGT
91401 TCGCGGGCCG GATGTCAGCC AGGCGTTCGG TCAGCAGATC GTGGAGCTGG
91451 TCGATCTGGG GGCCATGCGA GGC GTACCCG ACGTCGATGA CACGGGCGCG
91501 CAGACCACGT GCCTCCGCAT CGGCGACCAC GGCAGCCACA TGCTCCGGCG
91551 GCCCTGAAAT CACCGTAGAC GACGGCCCAT TGACCGCCGC GACGACCACG
91601 CCCGGCCGGT CACCGATCAG CTCAGCGGCC TGCTCGGCAC CGGTGCTCAG
91651 CGAGGCCATG TCACCGTGCC CTTGCAGCCG ACGAAGCGCG TCACTGCGTA
91701 CGGCTACGAT GCGCGCCGCA TCCTCCAGCG ACAGCGCCCC CGCGACGCAC
91751 GCGGCAGCCA TCTACCCTG CGAGTGCCCG ATCACAGCAG CCGGAGTGAC
91801 CCCGTAATCA GCCCACACCG CAGCCAGCGA GACCATCACC GCCACAACA
91851 CCGGCTGCAC GACCTCGACC CGGGACAGCT CACTCCCATC CCCGCGCAAC
91901 ACCGCACTCA GCGACCAGTC CACATACGCC GACAGCGCCC GCTCACACTC
91951 CGCAATCCGC GCCGCGAAGA CGGGGGACTC GTCCAGCAGC TGGGCACCCA
92001 TGCCACCCA CTGCGACCCC TGCCCCGGAA ACACCAACAC CGGCCCAGGA
92051 CCCACATCAC CAGCAACCCC GGCCACCACA CCCGCCGAAG CCTCACC CGC
92101 AGCCAACGCC CCCAGGCCAG CCGTCAACGC ATCGCGGTCA CGCCCCACCA
92151 CCACAGCCCG GTGCTCGAAC ACCGACCGGG TCGTGGTCAA CGACCAGCCC
92201 ACATCAGCCG CCGACGCATC CGCCGGCCCG GCCGCGAACT CGCCCAGCCG
92251 CCGCGCCTGT GCACGCAGCG CGTCCGGCGT CCGCCCGGAC ACCACCCACG
92301 GAACGACCCC ACCCGGCTCC TCGGCCACGG AGCCCGGCAC GTCTCTCTCC
92351 TCCGGTGGTG CCTCCTCCAG GATCAGATGC GCGTTCGTCC CCGAGAAGCC

92401 GAACGAGGAC ACCCCCGCCC GGCGCGGGCG CTCGCCCCGG GGCCACTTCA
92451 CCGGGTCGGT GAGCAGGCGC AGCCCGCTGC CGTCCCCTC CACGTGGGGC
92501 GAGGGGGCGT CGACGTGCAG GATGGCGGGC AGCAGGTCGT GCCGCAGGGC
92551 CAGGACCATC TTGATGACAC CGGCCACACC GGCGGCGATC TCGTGTGGC
92601 CGATGTTGGA CTTACCGCT CCCACCCACA GCGGCCGGTC CTCCGGCCGT
92651 TCCCGGCCGT AGGCGGAGAT GAGAGCCCCG GCCTCGATGG GGTCGCCGAG
92701 CGTGGTGCCG GTGCCGTGCG CCTCCACGGC GTCGATGTCC TCGGGGGCGA
92751 GGCGGGCGTT GGCGAGGGCG GCGCGGATGA CGCGTTCCTG GGCGGGGCCG
92801 TTGGGGGCGG TCAGGCCATT GCTCGCGCCG TCCTGGTTGA TCGCCGAACC
92851 CCGGATCACC GCGAGGACCT TGTGGCCCTT CCTGCGGGCG TCGGAGAGAC
92901 GCTCAAGGAG AACCACCCCC GTACCCTCCG CCATGCCCAT GCCGTCGCTG
92951 CTCGCCGAGA ACGGCTTGCA CCGTCCGTCG GGGGCCAGGC CGCGCAGTTC
93001 GCTGAAGCCG ATCAGCGGGG CGGGCGACGA CATCACGTAC GTGCCGCCCG
93051 CCAGCGCCAG CGAGCACTCC TGTGTGCGCA GGGCCTGGGT GCGAGGTGA
93101 AGGGAGACCA GCGACGAGGA GCACGCCGTG TCGACCGTCA CCGCGGGGCC
93151 TTCGAGGCC AGGGTGTAGG CGACGCGGCC GGAGGTGACG CTGCCGGAGT
93201 TGCCGATGGT GAAGTATCCG GCGGTGCCCT CGGGGACCTC GGACGCGCCG
93251 AGGGCGTAGT CGAGTCCGTC ACAGCCGATG AAGGTGCTGG TGTCGCTGGA
93301 GCGGAGGCTG AGGGGGTCGA TGCCGGCCCG TTCGATCGCC TCCCACGCCG
93351 TCTCCAGGGC GAGCCGCTGC TGCGGCGCCA TGGCCGCGGC CTCGGTGGGT
93401 CCGATGCCGA AGAAGGTGGG GTCGAAGTCA CCGGCGTCGT AGACGAAGCC
93451 GCCTTCCCGG ACGTAACTGG TGCCGGTGCT CTCGGGGTCC GGGTCGTAGA
93501 GGGAAATCGAG GTCCCAGTTG CGGTTGCCGG GCAGGGGCGC GACGGCGTCG
93551 CCGCCGGTGG AGACCAGCTC CCAGAACTCT TCGGGAGACC GGACTCCGCC
93601 GGGCAGCCGG CAGGCCATGC CGATGACCGC GACCGGTTCC TGGCCCGCCG
93651 ACTCGACGTC CTGCAGCCGG CGTTCCGTCT GACGAGGTC CGCGGTGACA
93701 CGCTTGAGGT ATTCCAGAAG TTTCTCTTCG GTGTGCGCCA TCCCGGTGAC
93751 AACCGCCCCT CTCCGCGAGA ACAGACCGCA GACTCGTCGA CGGCGCTAAA

93801 GCCCTCCTAA TACTCGGCTG TGTACCGCTC GCTGCCACGG GTGTCCGCAC
93851 TGGTCGGAGG CTCCGGCCCA GGGAACAGGG GCTTTCTTAG GGGCGCTTAA
93901 GCGGTGCCTG CCAGGGTGTG CCGGTGTCAG GCCGTCACGC CCTGATCAGC
93951 GGCCTCGCCC GTGCCGTGCC CGTGCGGTGCG GTGGGCCTGA CCGTCGGTCC
94001 GGACAACGCG AAGCGAGGCA TCGTGCCCAT CACGGATAGC AAGCCGGCCC
94051 CCACATTCCC CGACCTGGTC GACCCGTCGT TCTGGGCGCG GCCGCACGCG
94101 GAACGCGTGG CGCTGTTCGA GGAGATGCGC GGGCTGCCGC GGCCGGCGTT
94151 CATCCGGCAG AACATGCCCC GCGTGCCCTG GACGTTTCGGC TACCACGCGC
94201 TGGTCAAGTA CGCGGACATC GTGGAGGTGA GCCGCCGCCG GCAGGACTTC
94251 TCCTCGAACG GCGCGACCAC CATCATCGGT CTGCCGCCCG AGCTGGACGA
94301 GTACTACGGC TCGATGATCA ACATGGACAA CCCGGAACAC TCGCGGCTGC
94351 GGCGCATCGT CTCGCGTTCC TTCGGCCGCA ACATGATCCC CGAGTTTCGAG
94401 GCCGTGGCGA CCCGCACCGC CCGCCGCATC ATCGACGAGC TCATCGCGCG
94451 GGGACCCGGC GACTTCATCA GGCCCGTCGC CGCGGAGATG CCCATCGCCG
94501 TGCTCAGCGA CATGATGGGC ATCCCGGCGG AGGACCACGA CTTCTCTTTC
94551 GACCGGTCCA ACACGATCGT CGGCCCCCTC GACCCGGACT ACGTGCCGGA
94601 CCGGGCGGAC TCCGAACGCG CCGTGATCGA GGCCTCACGC GAACTCGGCG
94651 ACTACATCGC TGGCCTTCGT GCGGAACGGC TCGCCGCCCC CGGCAACGAC
94701 CTCATACCA AGCTCGTGCA AGTCCAGGCG GACGGCGAGC AGTTGACGCG
94751 GCAGGAACTC GTCTCCTTCT TCATCCTGCT CGTCATCGCC GGGATGGAGA
94801 CCACCCGCAA CGCCATCTCG CACGCGCTGG TACTGCTGAC CGAGCATCCC
94851 GAGCAGAAGC AGCTGCTGCT CTCGGACTTC GACACGCACG CGCCGAACGC
94901 GGTCGAGGAG ATCCTCAGGG TCTCCACGCC CATCAACTGG ATGCGGCGCG
94951 TCGCCACCCG CCACTGCGAC ATGAACGGCC ACAGGTTCCG CAGGGGCGAC
95001 CGGATCTTCC TGTTCCTACTG GTCGGGCAAC CGGGACGAAT CCGTCTTCCC
95051 TGACCCGTAC CGGTTTCGACA TCACGCGCGG GACGAACGCG CACGTCACGT
95101 TCGGCGCGGT GGGCCCGCAC GTCTGCCTCG GGGCCACCT CGCCCGTATG
95151 GAGATCACCG TCCTGTACCG GGAGCTGCTC GCGGCGCTGC CCCAGATCCA

95201 TGCCGTGGGG CAGCCCCGCA GGCTGGACTC CAGCTTCATC GAAGGGATCA
95251 AGCACCTGCA CTGCGCCTTC TGAGCACATA CGCTTCCCTC TGCGCATGTG
95301 CGCTCACGAC GCTCCGATCA GCGACTGCCA ACGACTGTCA GCGACCGGAC
95351 AGGGCCAAGG GCGGTGGGGA CATCAGGTGC ATGTCACCCG CGAGTATGGC
95401 CCGCTGCAGC TCCTGGAGCG GCGCCCCGGG TTCGAGCCCC AGCTCGTCGT
95451 TGAGCGTCTT GCGCACCGAC TGGTACACCT TCAGCGCGTC CGCCTGCCGC
95501 TCGGAGCGGT AGAGCGCCAG CATCAGCTGG CGGTAGAACG CCTCGCACAT
95551 CGGGTTCTCC GCGGTGAGGG CGTACAGCAT GCCCACGGCC TCGCGGTGCC
95601 GGCCGAGCTG GAGCTGGCAC TCGACGAGCA TCTCCTGACA CTCCAGGCGG
95651 ATCTCGGTCA GCCAGGTCGA GAAGCCGTCG ATGATCGGGC CGTTGGTGCC
95701 GGGACCGTTC CCGCCCTGCC CGAGGATCGG GCCGCGCCAC AGCGCGAGCG
95751 CCTGCCCCGAA ACAGGAGGCC GCCTCGTCGA ACCGCTTCTC CCTGAGCAAC
95801 GACCGCCCCA CGTCCACCAG TTCGGGGAAG ATCTGGGCAT CGATCTGGTC
95851 GTCGTCCCGC TTGTGCAGGA CGTACCCCGG CGCACGGGTC TCGACGGGGT
95901 TGCCCGCCGA ACCGGGCACC TTGAGGAACT TGCGGAGCTG GGAGATGTAC
95951 ACATGCAGTC CCGCCGTGGC GCGCCGCGGC AGGTCCTCGC CCCAGATCTC
96001 CCGCATCAGC TGCTCCAGGG AGACCACCCG GTCGGCGCGG ATGAGGAGCA
96051 CGGTGAGGAC GATCTCCACC TTCTGGGCGT TGATGGTGGC GTAGTCGTTT
96101 CCGTCCTTGA TGCGGAGCGG GCCCAGCATT TCGTATCTCA CCGAGCGTTC
96151 CCCCTTGCTG TCGCACGCTG CTGCGCACTG TCGGCCAGGG CCTTGGAGAT
96201 GACTTCCGTG ACGCCCTGCT GGTGCGTGTT CAGATAGAAG TGGCCCCCGG
96251 CGAAGACCTT GAGGTCGAAG GGGCCTTCCG TGTGCTGCTG CCAGGCCTCG
96301 ACCTCGTCCA GCGGCGCCTG CGGGTCCCCG TCGCCCACCA GGGCGGTGAT
96351 GGGGACAGGAC AGCGGCGGCG ACGGGTTCCA CCGGTACAGC TCGACCGCCC
96401 GGTAGTCGTT GCGGACGACC GGGATGATCT CCGCGAGCAG TTCCTCGTCG
96451 TCCAGGAACC GCGGGTCAGT GCCACCGGCC CGGCGCAGCT CGGCGGCCAA
96501 CTCGGTGTCG TCGAGGAGGT GTACGGTGCC GCGCCGGAAG CGGGACGGCG
96551 CGCGGCGTCC CGAGACGAAC AGCCGGCAGG GCTGCTTCCC CGTGCGCTCG

96601 CGGAGCCGCT GGGCGACTTC GTAGGCGAGG ACGGCGCCCA TGCTGTGGCC
96651 GAAGAACGCC AACGGGCGGT CGTCGAACGG GCCGAGCGCA TCGGTGATGA
96701 GGTCGGCGAG TTCCCCGATG TCGTCCAGGA GCCGCTCTCT GCGGCGGTCC
96751 TGTGCCCCGG GGTACTGCAC CGCGAGGACC TCGCTGTCGG TCGGGAGAGT
96801 GGGGGATTGC GCAAGGGGGT GGTAGTAGGA GGCCGAGCCG CCCGCGTGGG
96851 GGAAGCAGAC CAGGCGAACG ACGGCTTCCG GTCGGGGCCG GAAGCGACGT
96901 ATCCAAGGGT CCGACATATC GGGTGGGGGG AAGGCAGACA AGATCTTTCC
96951 CTTCGCCAGG AACGCTGACA ACGGTGTGTC GCCACATCAC ATAGCCGCTC
97001 CTGATCATGC GCAGCTCAAA GTTTAAACGG CAACGTCGCT AACGGGGGAG
97051 CAGGGCGGAA TCAGACATTC CCCATCCTTT ATTCCGCGAT TCTTACGTGA
97101 TCGAATCCCG GCGGCCAAGA TGGAGTAAAT TTCAATATGA ATGCTTAACG
97151 CCGCACAGCT TGTACGGCGG GCCGCCCGGG CGGTGACTGG CGTCCCTGCC
97201 AGCCGTGATG GCCTGACGAG GCCTCCGGGA TCCATCCCCC GCCCGCTGTG
97251 GCCGAGTTCT TTGCGGGATT ATTACGTTGC ATTGGTTTGC TTCGTGGCCC
97301 GGGCCGTTGG CCTGCGCTAT TTGGCAGCCT TCCGTCATGG GTGGTAAAAG
97351 ATCGCCTTTC CCCTCTGGGG TGCCGGTCGA GCTGGCCTCG ACCGCGATTG
97401 TGGCTTGTTG TTTTCTTGTG GCGCCGCGTG TGAAACAGCG GCAGTTGGCC
97451 ACTCGCTCTG ACAGGCTCCG GGGACGGGGT TGTCACCTTT TGGGGTGA
97501 GGCCTCGTTC AAGGCGTCCT GGCCCGTGGT GCATCCGCGA TCGTCGTGCC
97551 ATGGGTGAAG TGGGAAGGAG CACAGAACGA TGAGCGAGAG CATGGCGTGG
97601 CTGACGCGGG ACGTCCGCAA GGCCCGCAAG GAGGGCAGTG CGGGGACCGC
97651 GCGGCGCCGA GCCGACCGGC TGGCGGACCT GGTGCCCCAC GCCCGCTCGG
97701 CGTCGCCGTA CTACCGGGAG CTCTACCACG GCCTGCCCCA GCGGATCGAG
97751 GACCCGACGC TGCTGCCGGT GACGGACAAG AAGCAGCTGA TGGACCACTT
97801 CGACGACTGG CCGACGGACC GCGACATCAC CTTCGAGAAG GTCCGCGCGT
97851 TCACCGACGA CCCCAGCTG ATCGGGCGGC GCTTCCTCGG CCGCTATCTG
97901 GTGGCCACCA CGTCGGGCAC CAGCGGCAGG CGCGGCCTGT TCGTGCTCGA
97951 CGACCGGTAC ATGAACGTGT CCTCCGCCGT CTCCTCCCGG GTGCTCGCCT

98001 CCTGGGCTCGG CCCCTCGGC ATCGCCCGGG CCGTCGTCCA CGGCGGCCGC
98051 TTCGCCCAAC TCGTCGCCAC CGAGGGACAT TACGTCGGCT TCGCCGGATA
98101 CTCCCGCCTG CGCCAGGACG GCGGAGCGCG CAGCAAGCTC GTCCGCGCCT
98151 TCTCTGTGCA CGAGCCGATG TCACGTCTGG TCGCCGAAct CAACGAGTAC
98201 CGGCCCCGCT TCGTCATCGG CTACGCCAGT ACGATCATGC TGCTCACC GC
98251 CGAACAGGAA GCGGGCCGGC TGCACATCGA CCCGGTGCTG GTCGAGCCCC
98301 CGGGCGAGAC GATGACCGAG AGCGACACCG ACCGCATCGC TCGGGCGTTC
98351 GGCGCCAAGG TGCGCACGAT GTACAGCGCG ACCGAGTGCA CCTACCTCAG
98401 CCACGGCTGC GCCGAGGGCT GGTACCACGT CAACGACGAC TGGGCCGTGC
98451 TCGAACCGGT CGACGCCGAC CACCGGCCCA CCCC GCCGGG GGAGTTCTCG
98501 CACACCACCC TGATCAGCAA CTCGCCAAC CGCGTCCAGC CGTTCCTCCG
98551 CTACGACCTG GGCGACAGCG TCATGCTCCG CCCC GACCCC TGCCCCTGCG
98601 GCACCCCCTC GCCCGCGATC CGGGTCCAGG GCAGGTCGGG CGACATCCTC
98651 ACCTTCCCCT CGGGCCGGGG CGACGACGTC AGCCTCGCCC CGCTCGCCTT
98701 CAGCAGCCTC TTCGACCGCA TGCCCGGAGT CGAGCTCTTC CAGATCGAGC
98751 AGACCGCGCC GTCGACCCTG CGCGTCCGCG TGGTCCAGGC GCCCGGCGCC
98801 GACGCGGACC ACGTGTGGCA GCGGGCCCAC GACGGGCTGA CCCACCTCCT
98851 CGCCGACAAC AAGCTCGACA ACGTAACCGT CGAACGGGGC GAGGAGCCGC
98901 CGCGGCAGGC ATCCGGCGGC AAGTACCGGA CGATCATCCC GCTCGCCGCC
98951 TGAACGCTCG CCGACTAGCC GCGCGCCGCC TGAGCTGCTC TCACCGCGCG
99001 TACGGGCGCA GCGGAGGCTC CTCGTCGACC CACGGCTGGC TGTGGATCAG
99051 CAGCTCGATC GGGAAGTTCA GCAGGCCGGG CAGGGCGTCG ACGGCCTCCT
99101 GGCTGTTGAG CGGCATGACC GGCTTGCGC AGTGCGCGCG GTCGATGCGG
99151 CTCGTGTCGG CGGGACCGGG GTGCTCGATC GCATCGGCGA CCAGGTCGTA
99201 GCTGACCTGG TCGACGACCA TGGCGATGTG GGTCGGCCAC GGCCGACCCG
99251 GACAGATGTC CTGGACGCCG ATGCGGTGGG CGCCCGGCAG CGACGGCGCC
99301 TCCCCGTCGG CCACCACGGA CTCGTCGGCG TATGAATAGA TCGTGGTGTA
99351 CGACGGTCCT GCGGGCATGG GCGTGCCGTC GGCGCCAGG GCCTTCGACC

99401 AGTTCGAGTC GCGGGCGAAC TGCAGGACCG ACGCCGGGCA GCCC GCCACC
99451 TCGGCGATCG GGCGGCAGGG CGAGGCCAGC CGGGTCCCCT GGAACGGGGA
99501 GCCCAGGGTC ACCATGTCGT CGACCTTCCC CGGCAGGTCC GGCCAGAAGC
99551 GCAGGGCCCA CGCCGTGAGG AGGCCGCCCT GGCTGTGCCC GACGAGATCG
99601 ACCTTCCGGC CGGTGGCCTC CTGGATCGCG CGGGTCGCGT ACACCACGTA
99651 CTCGACGGAC TCCTGCATGT CACGGAGCCC GCGACCGGGA GAATCCACCC
99701 AACAGGACTG GTAGCCCTTC TTCTTCAACT CGGCCATGTA GTTCCAGGCG
99751 TAGTTCTCCT CGCCCTTGAG GCCGGTCCCG GGCACGAAGA GGACGGTCGG
99801 CTTGTCACCG GCGTCACGCA GGTCCCCCAG CTCCGTCCCG CAGTGCAGCG
99851 CCTTGCGGAG CTCGGCCGCC GGTATCTCCA ACGGGGGAGA GGAAACATCC
99901 GCCGCCGAAG CGGCGGAGGC CGGAAGCACG GTGGCGGCCA GCACGGCCGC
99951 CACGAGTCCG CCGAGCCATG AGGACAAGCG CACGGTGACC TCCACAGGAA
100001 CCTTCACGAG TGAGCGGAAA CTCCCTCCGG AGGGAGCACC TCATCGTGCG
100051 GCGGCGCCAC AGTAGCCGTC AACTGCCCCA CGGGGCTGAG TAGTTGACAG
100101 TTGGCCGGGC TCGGCCGGCG AAGCGCCCGG GCCCCGCCGC CCCGCGCCGT
100151 GCGCGAGGGG TCCGTGACCT GGGTGGACGG TCCGGTTGGA CATCCGGGG
100201 GAGCCTCTGG CATGGTCGCC CGTCCGTCCC CCTCAAGAAC CGAAGGGAGC
100251 GTCACGATCA CGATGATCGA AGTCAGCACG CGCAGCATGA AGGAAGCGGC
100301 TGCCGCCGAG CAGCTCCGCG CGGAGACCAC GACACTGGAC ATTCCAAAGG
100351 GTTTCGACCT GTGGACGGCC GACGAGATCG CGGAGTGGCT CGACGGCGTC
100401 GAGGACGACC CGGCAGTCTC CGACGCCGAC TTCTACGCGG CCCAGCAGCG
100451 GTGCGACGGG TCCTCGGCAC CGAGGGCACC TGACCCGCCG GCGGCCCTGC
100501 GCGGCCCTAC GTGTGCAGCG CCCCCTCCTC CTCCACATGC CCCTCCGGCT
100551 CCAGCTGGAT CGTCGAGTGG GCCACGTCGA AGTGGCCCCC GACACACCGC
100601 TGAAGGCGCC CCAGGAGCTC CCCGTACCCG CTCGCGAGAG CCTCCTCCGT
100651 GACCACCACG TGCGCGGTGA GCACCGGCAT CCCCAGAGTG ACCGTCCAGC
100701 CGTGACGATC GTGCACGGCG ACCACGCCCC GCTCCTCCAG CAGGTGCCGG
100751 CGCACCTCGC CGAGGTGCGAC GTCCTGCGGG GTCGCCTCCA GCAGGACGTG

100801 CAAGGAGTCC CGCAGCAGGC CGTACGCGCG CGGCACGATC AGCAGGCCGA
100851 TGACGATCGA CGCGATCGGG TCGGCGGCCT GCCACCCCGT GAGCAGGATG
100901 ACCAGGCCGC CCACGATCAC CGCGACCGAG CCGAGCGCGT CGCCCAGCAC
100951 CTCCAGGTAC GCGCCCCGCA GATTGAGGCT CTTCTCCTTG GCGTCCCGCA
101001 GCAGCCACAG GCCCACCAGG TTGGCGGCGA GCCCGCCCAG CGCGACCACG
101051 AACATCAGGC CGCCCTTCAC CTCCACCGGC TCGCTGAACC GGCCGATCGC
101101 CGACCACAGG ACCCAGGCGA AGATGACGAC CAGGAGCAGC GCGTTCAGGA
101151 CCGCGGAGAA GATCTCCACG CGGTAGAACC CAAAGGTGCG CCGCGGCGTC
101201 GCGCCCCGCT GGGCGAGGGT GATGGCACCG AGGGCCAGCG AGACGCCGAC
101251 CGCGTCGGTC AGGCTGTGCG CGGCGTCGGC GAGCAGCGCG AGGCTGCCGG
101301 ACAGGAGCGC GCCGACCACC TGGATGACGG TGATCGAGCC GCTGATGCCG
101351 ATGGTCCACA GCAGGCGCTT GCGGTACGTG CCGCTGAGAG TGCCGCCCGC
101401 CGCCCCGGCG GACGGACCGT GGTCGTGCCC CATGCCCGCG AGTGGACCAC
101451 GGCGGCGCGG CACCCGCCAC CGAGCGGCCG CCGGTCGGCT CAGTGCAGCC
101501 GGGCCTGGGT GGAGGTGTCG CGCTGGTGCG GGATGCCGAG CGGCGGCGGC
101551 AGCTCGCCCT GCTGCACCCT GACCGTGCGC ACGGGGGCGG GGACCCGGAT
101601 GCCCTCGGCG CGGTAACGCT GGTGCAGGCG CTTGATGAAC TCGTGCTTGA
101651 TGCGGTACTG GTCGCTGAAC TCGCCGACGC CGAGGATCAC CGTGAAGCTG
101701 ATCCGCGAGT CGCCGAAGGT GTGGAAGCGG ATCGCCGCCT CGTGGTCGGG
101751 GACCGCGCCG GTGATCTCGG CCATCACCTC GTCCACCACC TCGGTCGTGA
101801 CCTTCTCGAC CTGCTCCAGG TCGCTGTCGT AGCTGACCCC GACCTGCACC
101851 ATGATCGACA GCTCCTGCTC GGGGCGGCTG TAGTTGGTCA TGTTGGTGCC
101901 GGCGAGCTTC GCGTTGGGGA TGATGACGAG GTTGTGGAG AGCTGGCGGA
101951 CCGTGGTGTT GCGCCAGTTG ATGTCGACGA CGTAGCCCTC CTCCCCGCTG
102001 CTGAGCTGGA TGTAGTCGCC GGGCTGCACG GTCTTCGCGG CGAGGATGTG
102051 CACGCCCGCG AAGAGATTGG CGAGCGTGTC CTGCAGTGCG AGGGCGACCG
102101 CGAGACCTCC CACGCCGAGG GCGGTGAGCA GCGGTGCGAT GGAGATGCCG
102151 AGGGTCTGAA GGACGATGAG GAAGCCCATC GCGAGCACCA CGACGCGGGT

102201 GATGTTTCACG AAGATGGTGG CCGATCCGGC CACTCCGGAG CGGGACTGTG
102251 CCACGGCCTT CACCAGGCCG GTGACGATCC GGGCCGCCGT GAGCGTGGCG
102301 GCCAGGATGA GCAGCGCGGT CAGCGTCATG GTGACGTTGC GTCCGGTGCG
102351 CGGCGTGAGC GGCAGCGCGC CCGCCGCGGC GGCGAGCCCG GCGGTGATGG
102401 CCGCGCAGGG CACGAGGGTG CGCAGGGCGT CGACGATGAC GTCGTCACCG
102451 CTCCACCGGG TTTTGCTCGC CCGTTCGCCG AGCCACCTCA GAAGTGC GCG
102501 GAGCAGCAGC CCGGCGACGA CGCCGGCGAC GACCGCGATA CCGGCCACGA
102551 TCCAGTCGTG CAGTGTGAGG GCACGGGTCA TCAGTTCGCT CCCGTCGTAC
102601 GGGGGGAGTG CGCCTGTGTG GGGCGTATGT GATGTGACGT CACCTTGTGA
102651 TACCTGCTCG ATTCCGGGGA GTGCGGTCAC GCCGGGACGA GAGCTCGGTT
102701 CCGGCGCGGA CGTCATCCTG CCCCATCCGC CCACGGCAGG CGTGCATACC
102751 CCCACCTGGA TCTTCACAGA CCGGCCACGT CTGTCCATGC GCCGATGAGC
102801 GCGCTGCCCCG TGGTAAAGCA TTGAGTCAGG CGATTTGGCC ACTCGGCACT
102851 CGGCGGACCG GTCGAGCCGG TCGATCTACG TGAGCGGAGG CGGTTGAGCA
102901 TGGCGTCCAT GTGCAGACCC GGAATGTCAC CCGTCAATTC GCACAACGAG
102951 TGGGATCCGC TGGAGGAGAT CATCGTCGGG CGGCTGGAGG GCGCGACCAT
103001 TCCCTCCAGC CATCCGGTCG TGGCGTGCAA CATCCCGACC TGGGCGGCAC
103051 GGCTGCAGGG TCTCGCCGCC GGGTTCGAGT ATCCGCAGCG GCTGATCGAG
103101 CCGGCGCAGC AGGAGCTCGA CCAGTTCATC GCTCTCCTGC AATCCCTCGA
103151 CGTCACAGTG AGACGGCCGG CGGCCGTCGA CCACAAGCAC CGCTTCGGGA
103201 CCCCCGACTG GCAGTCGCGC GGCTTCTGCA ATTCCTGTCC GCGGGACAGC
103251 ATGCTCGTCG TCGGCGACGA GATCATCGAG ACCCCGATGG CGTGGCCGTG
103301 CCGCTGTTTC GAGACGCACT CGTACCGCGA ACTCCTCAAG GACTACTTCC
103351 GGCGCGGCGC GCGCTGGACG GCGGCGCCGC GCCCCAGCT CACCGAGGCC
103401 CTGTACGAGA AGGACTTCCG CCCTCCCGAG GAGGGCGAAC GATGCGCTAC
103451 ATCCTCACCG AGTTCGAGCC GGTGTTGAC GCGGCGGATT TCGTGCGGGC
103501 GGGCCGCGAC CTGTTCTGTA CGCGGAGCAA CGTCGCCAAC CTGCTGGGCA
103551 TCGAGTGGCT GCGCCGCCAC CTTGCGGGCCG GAGTACCGCG TGCCACGAGA

Table 1

positions_all

gene	function	start	end
gdhA	glutamate dehydrogenase (partial)	1038	0
dapA	dihydrodipicolinate synthase	2140	1220
orf3	putative transcriptional activator	2211	3152
orf4	hypothetical protein	3264	3680
orf5	hypothetical protein	4307	3684
orf6	hypothetical protein	4570	4758
orf7	hypothetical protein	5058	5612
acpX	acyl carrier protein	6010	5693
ksX	ketoacyl synthase	8531	6045
monCII	probable epoxihydrolase/cyclase	9542	8643
monE	methyltransferase	10426	9596
monT	monensin resistance gene (ABC-transporter)	10656	12191
monRII	probable repressor	12205	12780
monAIX	thioesterase	13829	13023
monAI	polyketide synthase loading & module 1	14121	23198
	KS-L	14172	15486
	AT-L malonate specific	15777	16880
	ACP-L	17019	17276
	KS1	17358	18626
	AT1 methylmalonate specific	18960	19976
	DH1 (potential)	20019	20519
	KR1 (inactive)	21636	22241
	ACP1	22536	22793
monAII	polyketide synthase module 2	23205	29921
	KS2	23307	24569
	AT2 methylmalonate specific	24891	25913
	DH2	25953	26369
	ER2	27600	28463
	KR2	28485	29042
	ACP2	29313	29570
monAIII	polyketide synthase modules 3 & 4	29974	42372
	KS3	30076	31347
	AT3 malonate specific	31798	32838
	DH3	32884	33465
	KR3	34692	35181
	ACP3	35553	35811
	KS4	35899	37170
	AT4 methylmalonate specific	37489	38511
	DH4	38557	38982
	ER4	40123	40986
	KR4	41005	41562
	ACP4	41848	42105
monAIV	polyketide synthase modules 5 & 6	42448	54564
	KS5	42628	43890
	AT5 ethylmalonate specific	44221	45243
	DH5	45289	45744
	KR5	46785	47337
	ACP5	47593	47850
	KS6	47947	49218

	AT6 malonate specific	49579	50601
	DH6	50644	51075
	ER6	52222	53102
	KR6	53101	53661
	ACP6	54052	54306
monAV	polyketide synthase modules 7 & 8	54614	66934
	KS7	54716	55978
	AT7 methylmalonate specific	56300	57319
	DH7	57358	57802
	KR7	59048	59608
	ACP7	59867	60124
	KS8	60185	61453
	AT8 malonate specific	61808	62839
	DH8	62882	63316
	EP8	64577	65437
	KR8	65456	66016
	ACP8	66404	66661
monAVI	polyketide synthase module 9	66952	72054
	KS9	67075	68340
	AT9 malonate specific	68698	69729
	KR9 (potential)	70735	71262
	ACP9	71536	71783
monH	probable regulator	72051	74993
monCI	FAD containing epoxidase	76541	75051
monBII	double bond isomerase	76960	76538
monBI	double bond isomerase	77450	77016
monAVIII	polyketide synthase modules 11 & 12	88708	77447
	KS11	88612	87344
	AT11 methylmalonate specific	87022	85993
	KR11	85111	84562
	ACP11	84292	84035
	KS12	83962	82694
	AT12 methylmalonate specific	82354	81335
	DH12 (potential) delta	81286	80855
	ER12 (potential)	79618	78914
	KR12	78895	78337
	ACP12	78070	77812
monAVII	polyketide synthase module 10	93741	88816
	KS10	93636	92368
	AT10 methylmalonate specific	92040	91021
	KR10	90132	89584
	ACP10	89322	89068
monD	P450 oxygenase	94081	95273
monRI	probable activator	96141	95338
monAX	thioesterase	96941	96138
orf29	cell wall biosynthesis capK homologue	97580	98953
lipB	lipase B	99983	98991
orf31	ion pump	101433	100507
orf32	membrane structural protein	102581	101490
amtA	glycine amidinotransferase	102924	103450

- ble II

GdhA, glutamate dehydrogenase (partial coding sequence) Length: 346 amino acids

```
1  LTTRPDTKTA LSQKTALSQ L TEIEHRNPA QPEFHQAARE VLETLAPVIA
51  ARPEYAEAGL IERLCEPERQ IVFRVPWQDD HGRVVRVNRGF RVEFNSALGP
101 YKGGLRFHPS VNLGVIFLGL FEQIFKNALT GLGIGGGKGG SDFDPRGRSD
151 AEVMRFCQSF MTELYRHIGE HTDVPAGDIG VGGREIGYLF GQYRRITNRW
201 EAGVLTGKGR NWGGSLIRPE ATGYGNVLFA AAMLRRERGET LEGRTAVVSG
251 SGNVAIYTIQ KLAALGANAV TCS DSSGYVV DEKGIDL DLL KQVKEVERAR
301 VDTYAQR RGA SARFVPGR RV WEVPADIALP SATQNELDAD DATA LI
```

DapA, dihydrodopicolinate synthase Length: 307 amino acids

```
1  MTLASSLEPT TEPLFNGLYV PLVTPFTDDL RLAPEALARL ADEALSAGAS
51  GLVALGTTAE AATLTAEERE TVIRVCSAAC RAHGAPLIVG VGTNDTATAI
101 TALRELAARG DVAAALVPAP PYIRPGEAGT LAHFAALAEH GGLPLVVYDI
151 PYRTGQTLGA GTITALGRLP EVVG I KHATG SIDPTTMELL DSPLPGFAVL
201 GGDDIVLSPL VAAGAHGGIV ASANLRTADY AEMIALWRRG SAAPARALGA
251 DLARLSAALF TEPNPTVIKG VLHAQNRIPS PAVRMPLLA SADSVRRAAP
301 LAASRK*
```

ORF3, putative transcriptional activator protein Length: 314 amino acids

```
1  MLDVRRRLHLL RELDRRG TIA AVAEALTFTA SAVSQQLGVL EREAGVPLLE
51  RSGRRVLTTP AGRSLVAHAD AVLNRLEQAV AELAGARDGI GGPLRIGTFP
101 SGGHTIVPGA LAELASRHPA LEPMVREIDS ARVSDGLRAG ELDVALVHDY
151 DFVPATPDTT VDEVPLLEEP MYLVTHAADT ATDSGSGSTL AALLGPCAEV
201 PWITARDGTT GHAMAVRACQ AAGFQPRIRH QVNDFRTVLA LVAAGQGAGF
251 VPRMAAEPSP AGVVLT KLPL FRRSKVAFRA GGGAHPAIAA FVAAATTAVE
```

01 RMAGSRGPAG GSE*

ORF4, hypothetical protein Length: 139 amino acids

1 MADDAYLFLL PDRHPRLGAA LAAVGALECT ETPAVHAWLQ AHEASVSSEQ
51 VRILPADAET LIPKDAERLP VPLSEEEALK VEQECAPQTV TDMESELLAF
101 RETTQDWQAL VHRALTAGIP AQRIARLTGL DPTEEIGRL*

ORF5, hypothetical protein Length: 208 amino acids

1 LAVAACAAVV LPIDAVVRIS AADVGVLVFF AYLLPYLAIT MTVFVSVAPE
51 QVRSWARREA RGTFLQRYVL GTAPGPGGSL FIAAAALVVA VLWLPGHLST
101 TFSALPRTL V ALALVVAWI CVVVAFAVTF QADNLVENER ALEFPGERSP
151 AWADYVYFAL AAMTTFGTTD VDVTSRDMRR TVAANTVIAF VFNTVTVAIL
201 VSALGGR*

ORF6, hypothetical protein Length: 63 amino acids

1 MTVMDKLKQM LKGHEDKAGQ GIDKAGDFVD GKTQGKYSQ VDTAQDKLRD
51 QFGSDQQEPP QR*

ORF7, hypothetical protein Length: 185 amino acids

1 MGTAQSQEQA AAPGACAAV RFVLCGGGVG LASSFAVVAL ASWVPWALAN
51 ALVAVVSTVV ATELHARFTF GAGGRATWRQ HAQSAGSAAA AYAVTCVAME
101 VLQQLVAAPG AVLEQVVYLS ASALAGVARF VVLRLVVFAR NRSLPAAAV
151 RTARPVRRVP APVPATVAHA ASRPAGPAAL CPAA*

AcpX, acyl carrier protein (ACP) Length: 106 amino acids

1 MTSTDHTSGQ DATELEKQLA AATPEEREKL LTDTIRTQAG TLLNTTSLDD
51 SNFLENGLNS LTALELTKTL MTLTGMEIAM VAIVENPTPA QLAHHLGQEL
101 AHTTA*

KsX, ketoacyl-ACP synthase Length: 829 amino acids

1 VANEEKLVEY LKWTTAELHQ AQQQLRELKA AQHEPIAVVS MACRLPGKTR
51 TPDDLWDLVS EGRDAVTGFP DDRAWELPEE RPYAELGGFL DDAAGFDAGF
101 FDISDTEAVA TEPLQRLMLH LAWETVERGH IAPHTLRSTL TGVYVGATGH
151 DYATRLETAP DELLPYLGGL TSGSLVSGRI AYALGLEGPA ISVDTACSSS
201 LVALHLACQA LRRGECGLAL AGGGTVMSTP HTFHAFHQK SLAQDGRCKP
251 FAAAADGMGL GEGVGLVLE RLGDARKNGH PVLAVIRGSA VNQDGAGYGL
301 AAPNGPSQQH VIRAAADAG LTPDQIDAVE AHGTGTPIGD AIEVQALLAT
351 YGADRSPDRP LWLGSVKSNT GHTQGAAGAA ALIKMVQAFR HGTLPPPTLHV
401 DRPTPLAAWK KGAVRLLTEA VDWPRREEPR RVGISAFATS GTNAHLILEE
451 PPVDEAPVPD AARDQTSPVA PELPVAWSLS ARTPEALRAQ AKALVTHLAA
501 TDPAPSPAEV AYSLAATRSP LEHRAVLTGT DHTELLAAAR ALAAGEDHPD
551 LVRSTPGAGP KKIAWHFDGR PADGVTTGAA PGAKPGATFG ATFGAAFGGA
601 EFHSAFPLFA SAFDEARALL DTHLPTPLPT PHSELARFAV HTALARLLE
651 TGVRPHTLTG DGVGHIAAAY AAGILTDDA CRLAAAHAAA AQAAEGEQPA
701 PPDAYEPVLK QLTFQRATLT LTSTAPADTP IASADYWHHH LTSPAPTAPP
751 TPETHLLHL GALSPEGTQT SAVSALLTAL ARLHTTGGTV DWTPLVRRTP
801 HPRTIDLPTY SFQATRYWLH DHTAHAHV*

MonCII, probable epoxyhydrolase/cyclase Length: 300 amino acids

1 VKNLRIPVSQ TVSLNVRYRP ADGPGAPGRP FLLHGMLSN ARMWDEVAAR
51 LAAAGHPAYA VDHRGHGESD TPPDGYDNAT VVTDLVAAVT ALDLGALVA
101 GHSWGAHLAL RLAAEHPDLV AGLALIDGGW YEFDGPVMRA FWERTADVVR
151 RAQQGTTSAA DMRAYLRATH PDWSPTSIEA RLADYRVGPD GLLIPRLTST
201 QVMSIVAGLQ REAPADWYPK VTPVVRLLPL IPAIPQLSDQ VRAWVAAAEA

251 ALEQVSVRWY PGSDHDLHAG APDEIAADLL LLARSCEAMP GGKAGVRPA*

MonE, S-adeonosylmethionine-dependent methyltransferase Length: 277 amino acids

1 VNKTVAPEPS DIGHYDHKV FDLMTQLGDG NLHYGYWFDG GEQQATFDEA
51 MVQMTDEMIR RLD PAPGDRV LDIGCGNGTP AMQLARARDV EVVGISVSAR
101 QVERGNRRAR EAGLADRVRF EQVDAMNLPF DDGSFDHCWA LESMLHMPDK
151 QQVLTEAHRV VKPGARMPA DMVYLNPDPS RPRTATVSDT TIYAALTDIG
201 DYPDIFRAAG WTVLELTDIT RETAKTYDGY VEWIRahrde YVDIIGVEGY
251 ELFLHNQAAL GKMPelgyIF ATAQRP*

MonT, putative monensin resistance gene (ABC-transporter) Length: 512 amino acids

1 MSADLGARRW WAVGALVLAS MVVGFDVTIL SLALPAMADD LGANNVELQW
51 FVTSYTLVFA AGMIPAGMLG DRFGRKKVLL TALVIFGIAS LACAYATSSG
101 TFIGARAVLG LGAALIMPTT LSLPVMFSD EERPKAIGAV AGAAMLAYPL
151 GPILGGYLLN HFWWGSVFLI NVPVVILAFL AVSAWLPESK AKEAKPFDIG
201 GLVFSSVGLA ALTYGVIQGG EKGWTDVTTL VPCIGGLLAL VLFVMWEKRV
251 ADPLVDLSLF RSARFTSGTM LGTVINFTMF GVLFTMPQYY QAVLGTAMG
301 SGFRLLPMTG GLLVGVTVAN KVAKALGPKT AVGIGFALLA AALFYGATTD
351 VSSGTGLAAA WTAAYGLGLG IALPTAMDAA LGALSEDSAG VGSGVNQSIR
401 TLGGSFGAAI LGSILNSGYR GKLDLDGVPE QAHGAVKDSV FGGLAVARAI
451 KSNGLADSVR SAYVHALDVV LVVSGGLGLL GVVLA VVWLP RHVGQSTAKT
501 AESEHEAADA V*

MonRII, probable repressor protein Length: 192 amino acids

1 VPGLRERKKA RTKAAIQREA VRLFREQGYT ATTIEQIAEA AEVAPSTVFR
51 YFATKQDLVF SHDYDLPFAM MVQAQSPDLT PIQAERQAIR SMLQDISEQE

101 LALQRERFVL ILSEPELWGA SLGNIGQTMQ IMSEQVAKRA GRDPRDPAVR
151 AYTGAVFGVM LQVSMDWAND PDMDFATTLD EALHYLEDLR P*

MonAIX, thioesterase Length: 269 amino acids

1 MDRGTAARAP QIGDEFGAAT GNGVWLRRYH AAAEAPVRLV CFPFAGGSAS
51 YYFGLSGLLA PGVEVLAVQY PGRQDRHAEP CLASVAELAD GVVPHLPCDG
101 KPFALFGHSL GAIVAFEVAR RLRGPAGPGL PVHLEFVSGGL ARPYPAGRS
151 GAFGDADILA HLRAMGGTDE RFFRSPELQE LVLPALRADY RAVATYEAPG
201 PGRLDCPITA LIGDADERTS PEQAATWRER TGAAFDLRVL PGGHFYLDGC
251 QEQVAADVTE ALTAGPGV*

MonAI, polyketide synthase multienzyme MONS 1, housing loading module and extension module 1 Length: 3026 amino acids

1 MAASASASPS GPSAGPDPIA VVGMACRLPG APDPDAFWRL LSEGRSAVST
51 APPERRRADS GLHGPGGYLD RIDGFDADFF HISPRAVAM DPQQRLLLEL
101 SWEALEDAGI RPPTLARSRT GVFVGAFWDD YTDVLNLRAP GAVTRHTMTG
151 VHRASILANRI SYAYHLAGPS LTVDTAQSSS LVAVHLACES IRSGDS DIAF
201 AGGVNLICSP RTTELAAARF GGLSAAGRCH TFDARADGFV RGE GGLVVL
251 KPLAAARRDG DTVYCVIRGS AVNSDGT TDG ITLPSGQAQQ DVVRLACRRA
301 RITPDQVQYV ELHGTGTPVG DPIEAAALGA ALGQDAARAV PLAVGSAKTN
351 VGHLEAAAGI VGLLKTALSI HHRRLAPSLN FTTPNPAIPL ADLGLTVQQD
401 LADWPRPEQP LIAGVSSF GM GGTNGHV VVA AAPDSVAVPE PVGVPERVEV
451 PEPVVVSEPV VVPTWPVSA HSASALRAQA GRLRTHLAH RPTPDAA RVG
501 HALATTRAPL AHRAVLLGGD TAELLGSLDA LAEGAETASI VRGEAYTEGR
551 TAF LFSGQGA QRLGMGRELY AVFPVFADAL DEAFALDVH LDRPLREIVL
601 GETDSGGNVS GENVIGEGAD HQALLDQTAY TQPALFAIET SLYRLAASFG

651 LKPDYVLGHS VGEIAAAHVA GVLSLPDASA LVATRGRIMQ AVRAPGAMAA
701 WQATADEAAE QLAGHERHVT VAAVNGPDSV VVSGDRATVD ELTAAWRGRG
751 RKAHHLKVSH AFHSPHMDPI LDELRAVAAG LTFHEPVIPV VSNVTGELVT
801 ATATGSGAGQ ADPEYWARHA REPVRFLSGV RGLCERGVTT FVELGPDAPL
851 SAMARDCFPA PADRSRPRPA AIATCRRGRD EVATFLRSLA QAYVRGADVD
901 FTRAYGATAT RRFPLPTYPF QRRHWPAAA GVGQQPETPE LPESSESSEQ
951 AGHEREEGAR AWGGPEGRLA GLSVNDQERV LLGLVTKHVA VVLGDASGTV
1001 QAARTFKQLG FDSMAAAELS ERLGTETGLP LPATLTFDYP TPLAVAAHLR
1051 AELTGTPAPA GSAPATGALG AGDLGTDEDP VAIVAMSCRY PGGAGTPEDL
1101 WRLVADGADA IGDFPTDRGW DLARLFHPDP DRSGTSCTRQ GGFLYDAADF
1151 DAEFFDISPR EALAVDPQQR LLECAWEAF ERAGLDPRAL KGSPTGVFVG
1201 MTGQDYGPRL HEPSQATDGY LLTGSTPSVA SGRLSFSFGL EGPALTVDTA
1251 CSSLVTLHL AAQALRRGEC DLALAGGATV LATPGMFTEF SRQRGLAPDG
1301 RCKPFAAGAD GTGWAEGVGL VLLERLSEAR RKGHAVLAVI RGSAINQDGA
1351 SNGLTAPNGP SQQRVIRAAL AAARLTADDEV DVVEAHGTGT TLGDPIEAQA
1401 LLATYGQGRS AERPLWLGSV KSNIGHTQAA AGVAGVIKMV MAMRHDLLPA
1451 TLHVDEPSGH VDWSTGAVRL LTEPVVWPRG ERPRRAAVSS FGISGTNAHL
1501 VLEEAGQDEY VAGAADDAGP VDGAVLPWVV SGRTGAALRE QARRLRELVT
1551 GGSADVSVSG VGRSLVTTRA VFEHRAVVVG RDRDTLIGGL EALAAGDASP
1601 DVVCGVAGDV GPGPVLVFPQ QGSQWVGMGA QLLGESAVFA ARIDACEQAL
1651 SPYVDWSLTE VLRGDGRELS RVDVVQPVLW AVMVSLAAVW ADHGVTAAV
1701 VGHSQGEIAA VVVAGALTLE DGAKIVALRS RALRQLSGGG AMASLGVGQE
1751 QAAELVEGHP GVGIAAVNGP SSTVISGPPE QVAAVVADAE ARELRGRVID
1801 VDYASHSPQV DAITDELTHT LSGVRPTTAP VAFYSAVTGT RIDTAGLDTD

1851 YWVTNLRRPV RFADAVTALL ADGHRVFIEA SSHPVLTGL QETFEEAGVD
1901 AVTVPTLRRE DGGRARLARS LAQAFGAGCA VRWENWFPAT GTSTVELPTY
1951 AFQRRRYWLE APTGTQDAAG LGLAAAGHPL LGAATEIADG DIRLLTGRIS
2001 RHSHPWLAQH TLFGAAVVPA SVLAEWALRA ADEAGCPRVD DLTLRTPPLV
2051 PETAGVQVQI VVG PADARDG HRDFHVIYARP DGKDASEGEG IAE GEGASEG
2101 EGASGGTDAP WTCHADGRLV AEPTGTASED SPDTVWPPPG AEPVDLGDFY
2151 ERAAATGVGY GPVFTGLRAL WRRDGELFAE AVL PQEAPET AGFGMHPALL
2201 DAALHPALLG ERPAEEDKVV LPFTLTGVTL WATGATSVRV RLTPLD DDDPD
2251 ASADGRAWRV GVS DPTGAEV LTCEALVAVA AGRRELRAAG ERVSDLYAVE
2301 WVPVPGPGPV GEGADFSGWA GLGECGERWE CVGRVERWYE DLDALGAAVE
2351 GGASVPSVVL ATAAAAPGGA GDGAADALSA VRWTGALLDQ WLADARFADA
2401 RLVVITSGAV ATGDDFLPDP AAAAVRGLVE QAQVRHPGRI LLVDTEAGAG
2451 LGVGAGVDDA LLEQAVAMAL GADEPQLALR AGRVLAPRLT APQDAAVTEA
2501 ARPLDPDGTV LITGPAGAPV ADLAEHLVRT GQCRHLLLLP GDGELEEMAE
2551 ELRGLGATVD LSTADPADPT ALAEVVAAVE GDHPLTGVIH ATGVVDAFDP
2601 GDSASDL MID SASDSFAEAW SSRAGVTAAL HTATAHLPLD LFAVLSPAGA
2651 DLGIARSAAA AGADAFSAAL ALRRHTTVTT DTTAPPRTTA PPRTTASPR
2701 TALSSSRTTG VALAYGPPTA PRPGIKGTAP GRIPVLLDAA RAHGGGSPLL
2751 GARLAARALA AESAAEGVAG LPAPLRALAV AAAAAGAPTR RTAADRKPPA
2801 DWPARLAPLS APEQLRLLID AVRTHAAAVL GRTDPEALRG DATFKQLGLD
2851 SLTAVELRNR LVEDTGLRLP TALVFRYPTP AAIAAHLRER LTSPSETTAT
2901 QRSGGQTPAA GQASSALAPG GSAAGPPAAD TVLSDLTRME NTL SVLAAQL
2951 PHTETGEITT RLEALLTRWK TTNATANDSG DGNGGDDDA ERLKAASADQ
3001 IFDFIDNELG VGHGTSRVTP TPKAG*

MonAII, polyketide synthase multienzyme MONS 2, housing extension module
2 Length: 2239 amino acids

1 MASEEQOLVEY LRRVTTELHD TRRRLVQEED RRQEPVALVG MACRFPGGVA
51 SPEDLWDLVA AGKDAIEDFP TDRGWDLEAL YDPDPAAYGT SYVRHGGFVD
101 DAGSFDADFF GISPREALAM DPQQRLMLET SWELFERAGI EPVSLKGSRT
151 GVIYAGVSSSED YMSQLPRIPE GFEGHATTGS LTSVISGRVA YNYGLEGPAV
201 TVDTACSASL VAIHLASQAL RQRECDLALA GGVLVLSSPL MFTEFCRQRG
251 LAPDGRCKPF AAAADGTGFS EGIGLLLLER LSDARRNGHK VLAIVIRGSV
301 NQDGASNGLT APNDAAQEQV IRAALDNARL TPSEVDAVEA HGTGTKLGDP
351 IEAGALLATY GQHRARPLLL GSLKSNIGHT HATAGVAGVI KTVMAIRNGL
401 LPATLHVEEL SPHVDWDAGA VEVVTEPTPW PETGHPRRAG VSAFGISGTN
451 AHLILEEAPP EEDVPAPVVV ESGGVVPWVV SGRTPEALRE QARRLGEFVA
501 GDTDALPNEV GWSLATTRSV FEHRAVVVGR DRDALTAGLG ALAAGEASAG
551 VVAGVAGDVG PGPVLVFPQG GAQWVGMGAQ LLDESAVFAA RIAECERALS
601 AHVDWSLSAV LRGDGSELSR VEVVQPVLWA VMVSLAAVWA DYGVTPAAVI
651 GHSQGEMAAA CVAGALSLED AARIVAVRSD ALRQLQGHGD MASLSTGAEQ
701 AAELIGDRPG VVVAAVNGPS STVISGPPEH VAAVVADAEA RGLRARVIDV
751 GYASHGPQID QLHDLLTERL ADIRPTNTDV AFYSTVTAER LTDTTALDTD
801 YWVTNLRQPV RFADTIEALL ADGYRLFIEA SAHPVLGLGM EETIEQADMP
851 ATVVP TLRD HGDTTQLTRA AAHAFTAGAD VDWRRWFPAD PAPRTIDLPT
901 YAFQRRRYWL ADTVKRDSGW DPAGSGHAQL PTAVALADGG VVLNGRVSAE
951 RGGWLGGHVV AGTVLVPGAA LVEWVLRAGD EAGCPSLEEL TLQAPLVLPE
1001 SGGLQVQVVV GAADEQGGRR DVHVYSRSEQ DASAVWQCHA VGELGRASVA
1051 RPVRQAGQWP PAGAEPVEVG GFYEGVAAAG YEYGPAFRGL RAMWRHGDDL

1101 LAEVELPEEA GSPAGFGIHP ALLDAALHPL LAQRSRDGAG AGAHGGQVLL
 1151 PFSWSGVSLW ASEATTVRVR LTGLGGGDDE TVSLTVTDPA GGPVVDVAEL
 1201 RLRST\$ARQV RGSAGPGADG LYELRWTPLP EPLPVPAPAN GRDVAADLSG
 1251 CAVLGELVAE PGPGIDLEG C PCYPGVGALA DNASPPSMIL APVHSDTTGG
 1301 DGLALTERVL RVIQDFLAAP SLEQKQTRLA FVTRGAADTG STTGGSAAAPA
 1351 EAVDPAVA AV WGLVRS AQSE NPGRFVLLDT DAPLDQASVA PLVDAVRS AV
 1401 EADEPQVALR GGRLLVPRWA RAGEPVELAG PAGARAWRLV GGDSGTLEAV
 1451 VAEACDDIVL RPLAPGQVRV AVHTAGVNFR DVLIALGMYP DPDALPGTEA
 1501 AGVVTEVGPG VTRLSVGDRV MGMMDGAFGP WAVADARMLA PVPPGWGTRQ
 1551 AAAAPAAFLT AWYGLVELAG LKAGERVLIH AATGGVGMAA VQIARHVGAE
 1601 VFATASPGKH AVLEEMGIDA AHRASSRDLA FEDAFRQATD GRGVDVVLNS
 1651 LTGELLDASL RLLGDGGRFV EMGKSDPRDP ELVALEHPGV SYEAFDLVAD
 1701 AGPERLGLML DRLGELFAGG SLVPLPVTAW PLGRAREALR HMSQARHTGK
 1751 LVLDVPAPLD PDGTVLVTGG TGTIGA AAVE HLARTGESKH LLIVSRSGPA
 1801 AHGAEELVSR IAEFGAEATF VAADVSEPDA VAALIEGIDP AHPLTG VVHA
 1851 AGVLDNALIG SQTTESLTRV WAAKAAAAQQ LHEATRESRL GLFVMFSSFA
 1901 STMGTPGQAN YSAANAYCDA LAALRRAEGL AGLSVAWGLW EATSGLTGTL
 1951 SAADRARIDR YGIRPTSAAR GCALLAAARA HGRPDLLAMD LDARVPAASD
 2001 APVPAVLRTL AAAGAPATAR PTAAAAADGA TDWSGRLAGL TEEARLELLT
 2051 ELVCTHAAGV LGHADAGAVQ VDAPFKELGF DSLTAVELRN RIAAATGLKL
 2101 PAALVFDYPQ ARVLA AHLAE RLVPEGAGAM GGVSGAEGVR DAYGAGGPGG
 2151 DMTAQVLLEV ARVEHTLSAA VPHGLDRAAV AARLEALLAR CTATTAATGA
 2201 AGAAVEGDGD SDGDGA VDQL ETATAEQVLD FIDNELGV*

MonAIII, polyketide synthase multienzyme MONS 3, housing extension modules 3 and 4 Length: 4133 amino acids

1 MVSEEKLV DY LKRVSADLHA TRQRLREAEE RGQEPVAVVE AACRYPGGIR
51 TPEDLWDLVA AGGNALGAFP DNRGWDLRRL FHPDPDHPGT TYAREGGFLH
101 DADLFDPEFF GISPREAAVL DPQQRLLEEC AWEALERAGI DPRSLQGSRT
151 GUYAGAALPG FGTPHIDPAA EGHLVTGSAP SVLSGRLAYT FGLEGPAVTI
201 DTACSSSLVA VHLLAAHALRQ RECDLALAGG VTMVMTTPYVF TEFSRQRGLA
251 ADGRCKPFAA AADGTAFSEG AGLLVLERLS DARRAGHRVL AVIRGSAVNQ
301 DGASNGLTAP NGPAQQRVIR AALAGARLSP AEVDAVEAHG TGTRLGDPPIE
351 ADALLATYGO ERHGGRPLWL GSVKSNIGHT QGAAGAAGLI KMVQALRHET
401 LPATLYADEP TPHADWESGA VRLLSAPVAW PRGEHGEHTR RAGISSFGIS
451 GTNAHLILEE APAADAEGAG GDGDGDGGGV RPVVVRVGATG PREEQGGQGG
501 QEQHQQQRQQ RQRSSMMPTP HLPWLLSARS PAALRAQADA LANHVAHADH
551 SIADIGGTLL RRTLFEHRAV VLGTDRDERA AALAALAAGR AHPALTRAAG
601 PARNGGTAFI FTGQGSQRPG MGRQLYDTFD VFAESLDETC ARLDPLLEQP
651 LKPVLFAPAD TAQAAVLHGT GMTQAALFAL EVALYRQVTS FGIAPSHLTG
701 HSVGEIAAAH VAGVFSLADA CTLVAARGRL MQALPAGGAM LAVQAAEDDV
751 LPLLAGQEER LSLAAVNGPT AVVVSGEAAA VGEVEKALRG RGLKTKRLNV
801 SHAFHSPLIE PMLDDFREVA RGLTFHAPTL PVVSNLTGRL ADAELMADAE
851 YWVRHVRRPV RFHDGLRALS EQGVVRYLEL GPDPVLATMV QDGLPAPAEG
901 EEPEPVVAAA LRSKHDEGRT LLGAVAALHT DGQPADLTAL FPADAGQVPL
951 PTYRFQRRRY WRVAPDAAAP ARAAGLQETG HPLLPAVIRQ ADGGILLAGR
1001 LSLRTHPWLA DHTIAGGVPL PATAFVELAL LAGRHAACDT IDDLTLETPL
1051 LLDDTGTGVG AAVGAGADAL VDAIEVQLAL GAPDGSGRRA LTVHSRPADD
1101 AADDGDAADA ADAAGRGGPG GSGDLGDPGD PGDLGDGGGS RGWRRHATGI

1151 LSAGPAAEPA APDAAPWPPA DATA LDVDAL YARLDAQGYS YGPAFRAVHA
1201 AWRHGDDLYA DVRLADEQRA EADAFALHPA LLDAALHAVD ELYRGSEGRG
1251 QEQGQGGQEP EQGRGDADAP VRLPFSFS DI RHHATGATRL WVRLSPQGDD
1301 RLRLSLTDGE GGQVATVDAL QLRLIPADRW RAARPTTAAP LYHLDWHELP
1351 LPEPAETDPA AHSWAVLG AH DAGLAPAAHY PDLAALKAAV EAGEPVPDIV
1401 FAPFPAQGTE TDVPAQVRAH ARHALELLRD WLTTEAFAAA RLVVLT TGAV
1451 TARPEDGPAD LATAPVWGLV RAAQAEQPDH VVLVDIDKDI DKDTDEETDQ
1501 ATDAGTASRH ALPAALAAAA AQAETQLALR AGTVLVPRLA VVPPRTDTPA
1551 LHATAPESTT DTV DSTGIAG AAESGGTVLI TGGTGGLGQA VARHLAAAHG
1601 ARHLLLVSRR GDAAEGVAEL RADLADDGVD VRVAACDITD RDALAGLLAD
1651 IPA AHPLTAV VHTAGVIDDS LITAMTPERL DAVLAPKADA AWHLHELTRD
1701 KDLSAFVLFS SGASVLGNGG QANYAAANTF LNTLAEHRRA AGLAATSVAV
1751 GLWESASGGM AARLG DADRA RIHRTGVTGL TDEQALALFD AALTAEHPTV
1801 LATRF DRAVL RGQAAARTLQ PALRGLVRTP RPTASAGAIG STAATGSATD
1851 ENAPSSWAAR LARLSAADRD RALNELIREQ IATVLAHPSP DTIELGRAFQ
1901 ELGFDSL TAL ELRNRLSTAT GIRLPATLVF DHPSPTALVR HLHSHLPDEA
1951 QHTSPTAPGA SAEGTAATAT GIDDDPIAIV GMACRYPGGV TSPEQLWQLV
2001 ATGTDAIGPF PEDRGWD TAG LFD PDPDQVG HSYTREGGFL YDAARFDAGF
2051 FGISP REAAA TDPQQRLLLE TAWQAFEHAG IDPAALRGTP CGVITGIMYD
2101 DYGSRFLARK PDGFEGRIMT GSTPSVASGR VAYTFGLEGP AITVDTACSS
2151 SLVAMHLAAQ ALRQGECELA LAGGV TVMAT PNTFVEFSRQ RGLAPDGRCK
2201 PFAAAADGTG WGE GAGLVVL ERLSDARRKG HRVLALLRGS AVNQDGASNG
2251 MTAPNGPSQE RVIR TALAGA GRGPEDIDVV EAHGTGTTLG DPIEAQALLA
2301 TYGQGRPEDR PLWLGSVKSN IGHTQAAAGV AGVIKMVMAL RHEQLPTTLH

351 ADEPTPHVQW DGGGVRLLTE PVPWSRGERT RRAGVSSFGI SGTNAHLILE
2401 EPPEEDLPEP VAAEPGGVVP WVVSGRTPDA LREQARRLGE FVVGAGDVSA
2451 AEVGWSLATT RSVFEHRVAV AGRDRDDLVA GMQALAAGET PTDVVSAGAA
2501 SSGAGPVLVF PGQGSQWVGM GAQLLDESPV FAARIAECEQ ALSAYVDWSL
2551 SDVLRGDGSE LSRVEVVQPV LWAVMVSLAA VWADYGVTPA AVVGHSQGEM
2601 AAACVAGALS LEDAARIVAV RSDALRQLQG HGDMA SLGTG AEQAAELIGD
2651 RPGVVVAVN GPSSTVISGP PEHVAAVVAE AEARGLRARV IDVGYASHGP
2701 QIDQLHDLT EGLADIRPAN TDVAFYSTVT AERLTDTTAL DTDYWVTNLR
2751 QPVRFADTIE ALLADGYRLF IEASAHPVLG LGMEETIEQA DIPATVVPTL
2801 RRDHGD TTQL TRAAAHAF TA GADVDWRRWF PADPTPRTVD LPTYAFQHQH
2851 YWLEEPSGLT GDAADLGMVA AGHPLL GACV ELAESDSYLF TGRLSRRAPS
2901 WLAEHVVAGT VLVPGAALVE WVL RAGDEAG CPTIEELTLQ APLVLPESGG
2951 LQVQVVVGAT DEQSGRRDVH VYSRSEQDAS AVWVCHAVGV VSSEMPEAAA
3001 ELSGQWPPAG AEAVDVEDFY ARAAEAGYAY GPAFQGLRAL WRHGTELF AE
3051 VVLPEQAGGH DGFGIHPALL DAALHPLMLL DRPADGQMWL PFAWSGVSLN
3101 ADRATHVRVR LSPRGEAAER DLRVVIADAT GAPVLTVDAL TLRAADPGRL
3151 GAAARGGVDG LYTVDWTPLP LPQPLPLPRT DAGGSADWVI LSDNSSAALA
201 DAVSSATAAG GGAPWALLAP VGGGSADDGL PVVRRTL SLV QEFLAAPELT
3251 ESRLVIVTRG AVATDADGDV AASAAAVWGL IRSAQSENPG RFVLLDVEEE
3301 HLHPDGGELP YAALRHAVEE LDEPQLALRS GKFLVPRMTP AAAPEELVPP
3351 VGTSGWRLGT SGTATLENLS VIDAPEAFAP LEPGQVRISV RAAGMNFRDV
3401 LIALGMYPDK GTFAGSEGAG HVTEVGPGVT HLSVGDRVMG LFEGAFAPLA
3451 VADARMVVPI PEGWSFQEAA AVPVVFLTAW YGLVDLGRLR AGESLLIHAG
3501 TGGVGMAATQ IARHLGAEVF ATASPAKHGV LDGMGIDA AH RASSRDLDFE

3551 ETLRAATGGR GMDVVLNSLA GEFTDASLRL LAEGGRMVDM GKTDKRDPDR
 3601 VAAEHAGAWY RAFDLVPHAG PDRIGEMLAE LGELFASGAL APLPVQTWPL
 3651 GRAREAFRFM SQAKHTGKLV LEIPPALDPD GTVLITGGTG VLAAAVAEHL
 3701 VREWGVRHLL LAGRRGSEAP GSSELAEELT ELGAEVTFAA ADVSDPDAVA
 3751 ELVGKTDPAH PLTGVIHAAG VLDDAVVTAQ TPESLARVWA AKATAAHLHL
 3801 EATREARLGL FLVFSSAAAT LGSPGQANYA AANAYCDALV RQRRAEGLAG
 3851 LSIGWGLWQT ASGMTGHLGE TDLARMKRTG FTPLTTEGGL ALLDAARAHG
 3901 RPHVVAVDLD ARAVAAQPAP SRPALLRALA AGATPGARTA RRTAAAGSVA
 3951 PAGGLADRLA GLPHPERRRL LLDLVRGNVA GVLGHSDDHA VRPDTSEKEL
 4001 GFDSLTADEL RNRLAAATGL KLPAALVFDY PESATLVDHL LERLSPDGAP
 4051 PPVKDAADPV LNDLGRIESS LDALALDADA RSRVTRRLNT LLSKLNGAAT
 4101 AGSPADVTDL DALDALDDVS DDEMFEFIDR EL*

MonAIV, polyketide synthase multienzyme MONS 4, housing extension modules 5 and 6 Length: 4039 amino acids

1 MSSAEESSPD VSGTGVSGTG ESATGTSSTE AKLRQYLKRV TVDLGQARRR
 51 LREVEERAQE PIAIVSMACR FPGDTRTPEA LWDLVAEGGD AIDDFPTNRG
 101 WDLESLYHPD PDHPGTSYVR RGGFLYDAPA FDASFFGISP REALAMDPQQ
 151 RVLMETAWQL LERAGIDPAS LKLSATGVYI GAGVLGFGGA QPDKTVEGHL
 201 LTGSALSVLS GRISFTLGLE GPSVSVDTAC SSSLVSMHLA AQALRQGECD
 251 LALAGGVTVM STPGAFTEFS RQGALSPDGR SKAFAASADG TGFSEGAGLL
 301 LLERLSDARR NGHKVLAIR GSAVNQDGAS NGLTAPNGPS QERVIRAALA
 351 NAGLGAAEVD AVEAHGTGTK LGDPIEAGAL LATYGRDRDE DRPLWLGSVK
 401 SNIGHPOGAA GVAGVIKVM ALQRELLPAT LYVDEPTPHV DWSSGSVRLI
 451 TEPVPWTRGE RPRRAGVSAF GMSGTNAHVI LEEAPPEEAA AAETPAEGTG
 501 AVVPWVSGR GEEALRAQAA QLAEHVRDDD QRPASPLEVG WSLATTRSVF

551 ENRAVVVGDD RDALLDGLRS LAAGEASPDV VSGAVGPTGP GPVMVFPGQG
601 GQWVGMGARL LDESPVFAAR IAECEQALSA YVDWSLTDVL RGDGSELARI
651 DVVQPV LWAV MVALAAVWAD QGIEPAAVVG HSQGEIAAAC VVGAISLDEA
701 ARIVAVRSVL LRQLSGRGGM ASLGMGQEQ AADLIDGHPGV VVA AVNGPSS
751 TVISGPPEGI AAVVADAQER GLRARAVASD VAGHGPQLDA ILDQLTEGLA
801 GIRPAATDVA FYSTVTAGHL TDTTELDTAY WVRNVRRTVR FADTIDALLA
851 DGYRLFIEVS PHPVLNLALE GLIERAAVPA TVVPTLRRDH GDTTQLARAA
901 AHAFAGADV DWRRWF PADP APRTVDLPTY AFQRQDFWPA PAGGRSGDPA
951 GLGLAASGHP LLGASVGLAS GDVHLLSGRV SRQSAAWLDD HVVAGQALVP
1001 GAAQVEWVLR AGDDAGCSAL EELTLQTPLV LPDTGGLRIQ VVVEAADAHG
1051 RRDVRLFSRP DDDDAFASTH PWTCHATGVL APAPTDGTNG TRDAADTLDG
1101 AWPPADAEPV PADDLYAQAD RTGYGYGPAF RGVRLWRHG KDVLA EVTLP
1151 KEAGDPDGFG IHPALLDAVL QPAALLPPT DAEQVWLPFA WNDVALHAVR
1201 ATTVRVRLTP LGERIDQGLR ITVADAVGAP VLTVRDLRSR PTDTGRLAAA
1251 ATRDRHGLFD LEWIAPENAA ENAAGPARDA SEGWVTLGED AASLADLLAS
1301 VEAGAPAPQL VAAPVEPDRT DDGLALATHV LDLVQTWLAS PLHDSRLVLV
1351 TRGAVTDADV DVAAA VWGL VRS AQSEHPG RFTLIDLGP DTLAAAMQAA
1401 HLEEPQLAVH GGEIRVPRLV RATTDPTAPN GTPEADRTAD PSEGLHRNGT
1451 VLITGGTGVL GRLVAEHLVT EWGVRHLLLA SRRGDQAPGS AELRARLSEL
1501 GASVEIAPAD VGDAEAVAAL IASVDP AHPL TGVIHAAGVL DDAVITAQTP
1551 ESLARVWATK ATAARHLHEA TRETPLDEFFV VFSSAAASLG SPGQANYAAA
1601 NAYCDALVQH RRAQGLAGLS IAWGLWQATS GMTGQLSETD LARMKRTGFA
1651 ALTDEGLLAL LDAARAH DRA YVVAADLDPR AVTDGLSPLL RALTAPATRR
1701 RVASEGLADG ALATRLAGLD ADGRLRL LTD VVREYVA AVL GHGSAARVGV

1751 DIAFKDLGFD SLTAVELRNR LSAACDVRLP ATLIFDHPTP QALATHLVDR
1801 LAGSTSATTT VNATAPAAAH VAAGADVDDAD TDDPVAIVAM TCRFPGGVAS
1851 PDDLWDL LDA RKDAMGAFPT DRGWDLERLF HPDPDHPGTS YTDQGGFLPD
1901 AGDFDAAFFG INPREALAMD PQQRLLLEAS WEVLERAGID PTTLKGTPTG
1951 TYVGLMYHDY AKSFPTADAQ LEGYSYLAST GSMVSGRVAY TLGLEGPAVT
2001 VDTACSSSLV SIHLATQALR HGECDLALAG GVTVMADPDM FAGFSRQRGL
2051 SPDGRCKAYA AAADGVGFSE GVGVLLLERL SDARRHGRRV LGVVRGSAVN
2101 QDGASNGLTA PNGPSQERVI RQALASGGLS SVDVDVVEGH GTGTTLGDPI
2151 EAQALLATYG QGRPEDRPLW LGSVKSNIH TQAAAGVAGV IKMVMAMRHG
2201 VVPASLHVDV PSPHVEWDSG AVRLAVESVP WPQVEGRPRR AGVSSFGASG
2251 TNAHVIVESV PDGLEEDSVS VGGEALETET DGRLVPWVVS ARSPQALRDQ
2301 ALRLRDFASD ASFRAPLADV GWSLLKTRAL HEHRAVVVGA ERAELIAALE
2351 ALATGEPHAA LVGPACSQAR VGGDDVVWLF SGQGSQLVGM GAGLYERFPV
2401 FAAAFDEVCG LLEGPLGVEA GGLREVVRG PRERLDHTVW AQAGLFALQV
2451 GLARLWESVG VRPDVVLGHS IGEIAAAHVA GVFDLADACR VVGARARLMG
2501 GLPEGGAMCA VQATPAELAA DVDGSAVSVA AVNTPDSTVI SGPSDEV DRI
2551 AGVWRERGRK TKALSVSHAF HSALMEPMLA EFTEAIRGVK FRQPSIPLMS
2601 NVSGERAGEE ITDPEYWARH VRNAVLFQPA IAQVADSAGV FVELGPAPVL
2651 TTAAQHTLDE SDSQESVLVA SLAGERPEES AFVEAMARLH TAGVAVDWSV
2701 LFAGDRVPGL VELPTYAFQR ERFWLSGRSG GGDAATLGLV AAGHPLLGA
2751 VEFADRGGCL LTGRLSRSGV SWLADHVVAG AVLVPGAALV EWALRAGDEV
2801 GCVTVEELML QAPLVVPEAS GLRVQVVVEE AGEDGRRGVQ IYSRPDADAV
2851 GGDDSWICHA TGVLSPE SAR LDTELGGVWP PAGAEPLDVD GFYAQAGEAG
2901 YGYGPAFRGL RAVWRHGQDL LAEVLPEAA GAHDGYGIHP ALLDATLHPL

2951 LAARFMDGSE DDQLYVPFGW AGVSLRAVGA TTVRVRLRPV GESVDQGLSV
 3001 TVTDATGGPV LSVDSLQTRP VKPSQLAAQ QPDVRGLFTV EWTPLPQTDA
 3051 DGEADWVVL S DGVGRLADV SAAGGEAPWA VVAPVDASVG DGREGLDGRL
 3101 VVERVLSLVQ EFLALPELAE S RLLVVTRGA VATGVDGDGD VDASAAAVWG
 3151 LVRSQAQSEN GRFILLDVDG DGDDQGPDLN GRHLPATLR HAAEELDEPQ
 3201 LALREGTLYV PRLTQARQSA ELVVPPGEPA WRLRMVHDGS LDALAAVACP
 3251 EALEPLAPGQ VRIAVHAAGI NFRDVLVALG MVPAYGAMGG EGAGVVTEVG
 3301 PEVTHVSVGD RVMGVFEGAF GPVVIAEARM VTPVPQGWD M REAAGIPAAF
 3351 LTAWYGLVEL AGLKAGERVL VHAATGGVGM AAVQIARHVG AEFATASPG
 3401 KHAVLEEMGI DAAHRASSRD LAFEGTFREA TGGRGMDVVL NSLAGEFIDA
 3451 SLRLLGDGGR FLEMGKT DVR AAEEVAAEHA DVSYTAYDLV GDAGPDRI SN
 3501 MLDKLVLEFA SERLKPLPVR SWPLDKAQEA FRFMSQAKHT GKLVL EIPPA
 3551 LDPEGTVLVT GGTGALGQVV AEHLVREWGV RHLLLASRRG PEAPGSDELA
 3601 SKLTGLGAEV TIVAADVSDP ASVVELVGKT DPSHPLTG VV HAAGVLEDGV
 3651 VTAQTPEGLA RVWAAKAAAA ANLHEATREM RLGLFVVFSS AAATLGSPGQ
 3701 ANYAAANAYC DALMQHRRAV GQVGLSVGWG LWEAPDAKPG VAADAKASAA
 3751 TVGKASALSD GTNGSAPQDT TGTAPQGMTG GLTDTDVARM ARIGVKGMSN
 3801 AHGLALFDAA HRHGRPHLVG FNLDLRTLAT HPLHTRPALL RGLATPTAGG
 3851 ASRPTATAGG QPADLAGRLA ALSPSDRHHT LVRLIREQAA TVLGHH PDSL
 3901 TTGSTFKELG FDSLTAVELR NRLSAATGLR LPAGLVFDHP DADILAEHLG
 3951 AQLAPDGDTP AGAEATDPVL RDLAKLENAL SSTLVEHLDA DAVTARLEAL
 4001 LSNWKAASAA PGSGSTKEQL QVATTDQVLD FIDKELGV*

**MonAV, polyketide synthase multienzyme MONS 5, housing extension
 modules 7 and 8 Length: 4107 amino acids**

1 MASEEELVDY LKRVAELHD TRQRLREVED RRQEPVAVVG MACRFPGGIE
51 TPEGLWELVA AGDDAIEPFP TDRGWDLEGI YHPDPDHPGT CYVREGGFLA
101 APDRFDSDFE GFSPREALAS SPQLRLLET SWEALERAGI NPASLKGSPT
151 GVVYGAATTG NQTQGDPPGK ATEGYAGTAP SVLSGRLSFT LGLEGPVTV
201 ETACSSSLVA MHLAANALRQ GECDLALAGG VTMSTPEVF TGFSRQRGLA
251 PDGRCKPFAA AADGTGWGEG AGLILLERLS DARRKGHKVL AVIRGSAINQ
301 DGASNGFTAP NGPSQRRVIR QALSSAHLST SEIDVVEAHG TGTRLGDPIE
351 AEALIATYGK EREDDRPLWL GSVKSNIGHT QAAAGVAGVI KVMALQREL
401 LPATLNVDEP TPHVQWEGGG VRLLETPVPW SRGERPRRAG ISSFGISGTN
451 AHVVLEEAPP EEDVPGPVAA EPEGVVPWV SARTEEALSE QARRLGEFVA
501 DTDPSADV WSLTTSRAIL EHRAVVVGRD RDALTAGLAA LAAGEESADV
551 VAGVAGDVGP GPVLVFPQG SQWVGMAQL LDESPVFAAR IAECEQALSA
601 YVDWSLSAVL RGDGSELSRV EVVQPVWAV MVSLAAVWAD YGVTPAAVIG
651 HSQGEMAAAC VAGALSLEDA ARVVAVRSDA LRQLMGQGM ASLGASSEQA
701 AELIGDRPGV CIAAVNGPSS TVISGPPEHV AAVVADAEER GLRARVIDVG
751 YASHGPQIDQ LHDLLTDRLA DIRPATTDVA FYSTVTAERL TDTTALDIDY
801 WVTNLRQPVR FADTIDALLA DGYRLFIEAS AHPVLGLGME ETIEQADIPA
851 TVVPTLRRDH GDTTQLTRAA AHAFATAGTV DWRRWFADP TPRTIDLPTY
901 AFQRRSYWLP VDGVDVRSR GLRRVEHSLP PAALGLADGA LVLTGRLAAS
951 GGGGGWLADH AVAGTTLVPG AALVEWALRA ADEAGCPSLE ELTLQAPLVL
1001 PGSGGLQVQV VVGPDGQGG RREVRVFSRV DSDDEAAGQD EGWSCHATGV
1051 LSPEPGAVPD GLSGQWPPTG AEPLISDLY EQAASAGYEY GPSFRGLRSV
1101 WRHGHNLLAE VELPEQAGAH DDFGIHPVLL DAALHPALLL DQNAPEGEEQE
1151 PAQPALRLPF VWNGVSLWAT GAATVRVRLA PHGGGETDDS AGLRVTVADA

201 TGAPVLSVDS LALRPADPEL LRTAGRAGSG TNGLEFTVEWT ALPPADVADH
1251 AAGDGWAVLG QDVPDWAGAD MPRHPDMASL SAALDEGTQA PAAVFEVETTA
1301 TSHATPNTAA DVTLDASGRA VAERTLHLLR DWLAEPRLA ETRLVLITHHA
1351 VTTPADDDVN AAPLDVPAAA LWGLIRSAQA EHPDRFVLLD TDAKANTDPG
1401 PDTSTDHSTA SGTYRTVIAR ALATGEPQLA VRAGELLAPR LARAATPTPE
1451 TPTPETQPD T GSGSEAGAGS GSGPGATLDP DGTVLIAGGT GMMGGLVAEH
1501 LVRAWSVRHL LLVSRQGPDA PDARDLADRL VGLGATVRIV AADLTDGRAT
1551 ADLVASVDPA HPLTGVIHAA GVLDDAVVTA QTS DQLARVW AAKASVAANL
1601 DAATSELPLG LFLMFSSAAG VLG NAGQAGY AAANAFVDAL VGRRRATGLP
1651 GLSIAWGLWA RGSAMTRHLD DADLARLRAG GVKPLLDEQG LALLDAARAT
1701 AAHTSLVVAA GIDVRGLNRD DVPAILRDLA GRTRRRAAAD STVDQAALER
1751 RLTGLDEAER RAVVTDVVRE CVA AVLGHRS AADV RTEANF KDLGFDSLTA
1801 VQLRNRLSAA SGLRLPATLA FDHPTPQALA AYLGTRL SGR TATPVAPVAP
1851 SAAATDEPVA IVAMACKYPG GATSPEGLWD LVAEGVDAVG AFPTGRGWDL
1901 ERLFHPDPDH PGTSYADEGA FLPDAGDFDA AFFGINPREA LAMPDQQRLL
1951 LEASWEVLER AGIDPTTLKG TPTGTYVGVM YHDYAAGLAQ DAQLEGYSML
2001 AGSGSVVSGR VAYTLGLEGP AVTVDTACSS SLVSIHLAAQ ALRQGECTLA
051 LAGGVTVMAT PEVFTGFSRQ RGLAPDGRCK PFAAAADGTG WGEGVGVLLL
2101 ERLSDARRHG RRVLG VVRGS AVNQDGASNG LTAPNGPSQE RVIRQALASG
2151 GLSSVDVDVV EGHGTGTTLG DPIEAQALLA TYGQGRP VDR PLWLGSVKSN
2201 IGHTQAAAGV AGVIKMVMAM RHGVVPASLH VDVPSPHVEW DSGAVRLAVE
2251 SVPWPEVEGR PRRAGVSSFG ASGTNAHVIV ESVPDGLGED SVSVSGEAPE
2301 TETDGR LVPW VVSARSPQAL RDQALRLRDA VAADSTVSVQ DVGWSLLKTR
2351 ALFEQRAVVV GRERAELLSG LAVLAAGEEH PAVTRSREDG VAASGAVVWL

2401 FSGQGSQVLG MGAGLYERFP VFAAAFDEVC GLLEGPLGVE AGGLREVVFR
2451 GPRERLDHTM WAQAGLFALQ VGLARLWESV GVRPDVVLGH SIGEIAAAHV
2501 AGVFDLADAC RVVGARARLM GGLPEGGAMC AVQATPAELA ADVDDSGVSV
2551 AAVNTPDSTV ISGPSGEVDR IAGVWRERGR KTKALSVSHA FHSALMEPML
2601 AEFTEAIREV KFTRPKVSLI SNVSGLEAGE EIASPEYWAR HVRQTVLFPQ
2651 GIAQVASTAG VFVELGPGPV LTAAQHTLD DVTDRHGPEP VLVSSLAGER
2701 PEESAFVEAM ARLHTAGVAV DWSVLFAGDR VPGLVELPTY AFQRRERFWLS
2751 GRSGGGDAAT LGLVAAGHPL LGAAVEFADR GGCLLTGRLS RSGVSWLADH
2801 VVAGAVLVPG AALVEWALRA GDEVGCVTVE ELMLQAPLVV PEASGLRVQV
2851 VVEEAGEDGR RGVQIYSRPD ADAVSGDDSW ICHATGTLTP QHTDAPNDGL
2901 AGAWPAAGAV PVDLAGFYER VADAGYAYGP GFQGLRAVWR HGQDLLAEVV
2951 LPEAAGAHDG YGIHPALLDA TLHPALLLDW PGEVQDDDGK VWLPFTWNQV
3001 SLRAAGAATV RVRLSPGEHD EAEREVQVLV ADATGTDVLS VGSVTLRPAD
3051 IRQLQAVPGH DDGLFSVDWT PLPLSRTDVS QTDADGDADW VVLSDBGVSL
3101 ADVVSAAGGE APWAVVAPVG ASAGGGLAGF DRREGLDGRL VVERVLSLVQ
3151 EFLAAPELAE SRLLVLTRGA VATGGDGDGD VDASAAAVWG LVRSQAQSEN
3201 GRFILLDMDV DVDVDVMDV DVDVDVDVDV DGDGNGSDLD PDLNGRRLPH
3251 ATLRHAAEEL DEPQLALRDG QLLVPRLVRA TGGGLVVAPT DRAWRLDKGS
3301 AETLESVAPV AYPGVMEPLG PGQVRLGIHA AGINFRDVLV SLGMVPGQVG
3351 LGGEGAGVVT ETGPDVTHLS VGDRVMGVLH GSFGPTAVAD TRMVAPVPQG
3401 WDMRQAAAMP VAYLTAWYGL VELAGLKAGE RVLIHAATGG VGMAAVQIAR
3451 HLGAEVFATA SAAKHVVLEE MGIDAHRAS SRDLAFEDTF RQATDGRGMD
3501 VVLNSLTGEF IDASLRLLDG GGRFLEMGT DVRTPEEVAA EYPGVTYTVY
3551 DLVTDAGPDR IAVMMSELGE RFASGALDPL PVRSWPLDKA REAFRFMSQA

3601 KHTGKLVLDV PAPLDPDGTV LITGGTGALG QVVAEHLVRE WGVRLHLLAS
3651 RRGLDAPGSG ELADRLSDLG AEVTVAAADV SDPASVVELV GKTDPSHPLT
3701 GVVHAAGVLE DGIVTAQTPE GLARVWAAKA AAAANLHEAT REMRLGLFVV
3751 FSSAAATLGS PGQANYAAAN AYCDALMQRR RAAGQVGLSV GWGLWEAPDA
3801 KPGVAADAKP DVAADAKTGV AADGTPQGMT GTLSGTDVAR MARIGVKAMT
3851 SAHGLALLDA AHRHGRPHLV AVDL DTRVLA HKPAPALPAL LRAFAGDQGG
3901 QGGGRGGGRG GGPARPAAAT TRQNVDWAAK LSVLTAEEQH RTLLDLVRTH
3951 AA AVLGHAGT DAVRADAAFQ DLGFDSLTA V ELRNRLSAST GLRLPATFIF
4001 RHPTPSAIAD ELRAQLAPAG ADPAAPLFGE LDKLETVITG HAHDESTRTR
4051 LAARLQNLLW RLDDTSARSD HAAGASDADG DAVENRDLES ASDDEL FELI
4101 DRELPS*

**MonAVI, polyketide synthase multienzyme MONS 6, housing extension
module 9 Length: 1701 amino acids**

1 MPGTNDMPGT EDKLRHYLKR VTADLGQTRQ RLRDVEERQR EPIAIVAMAC
51 RYPGGVASPE QLWDLVASRG DAIEEFPADR GWDVAGLYHP DPDHPGTTYV
101 REAGFLRDAA RFDADFFGIN PREALAADPQ QRVLLEVSWE LFERAGIDPA
151 TLKDTLTGVY AGVSSQDHMS GSRVPPEVEG YATTGTLSSV ISGRIAYTFG
201 LEGPAVTLDT ACSASLVAIH LACQALRQGD CGLAVAGGVT VLSTPTAFVE
251 FSRQRGLAPD GRCKPFAEAA DGTGFSEGVG LILLERLSDA RRNGHQVLGV
301 VRGSAVNQDG ASNGLTAPND VAQERVIRQA LTNARVTPDA VDAVEAHGTG
351 TTLGDPIEGN ALLATYGKDR PADRPLWLGS VKSNIGHTQA AAGVAGVIKM
401 VMAMRHGELP ASLHIDRPTP HVDWEGGGVR LLTDPVPWPR ADRPRRAGVS
451 SFGISGTNAH LIVEQAPAPP DTADDAPEGA ATPGASDGLV VPWVVSARSP
501 QALRDQALRL RDFAGDASRA PLTDVGWSSL RSRALFEQRA VVAGRERAEL
551 LAGLAALAAG EEHPAVTRSR EEAAVAASGD VVWLFSGQGS QLVGMGAGLY

601 ERFPVFAAAF DEVCGLLEGE LGVGSGGLRE VVFWGPRERL DHTVWAQAGL
651 FALQVGLARL WESVGVRPDV VLGH SIGEIA AAHVAGVFDL ADACRVVGAR
701 ARLMGGLPEG GAMCAVQATP AELAADV DGS SVSVA AVNTP DSTVISGPSG
751 EVDRIAGVWR ERGRKTKALS VSHAFHSALM EPMLGEFTEA IRGVKFRQPS
801 IPLMSNVSGE RAGEEITSPE YWARHVRQTV LFQPGVAQVA AEARAFVELG
851 PGPVLTAAQ HTLDHITEPE GPEPVVTASL HPDRPDDVAF AHAMADLHVA
901 GISVDWSAYF PDDPAPRTVD LPTYAFQGRR FWLADIAAPE AVSSTDGEEA
951 GFWAAVEGAD FQALCDTLHL KDDEHRAALE TVFPALSAWR RERRERSIVD
1001 AWRYRVDWRR VELPTVPGA GTGPDADTGL GAWLIVAPTH GSGTWPQACA
1051 RALEEAGAPV RIVEAGPHAD RADMADLVQA WRASCADDTT QLGGVLSLLA
1101 LAEAPATSSD TTSTSTSCG TGSLASHGLT GTLTLLHGLL DAGVEAPLWC
1151 ATRGAVSCGD ADPLVSPSQA PVWGLGRVAA LEHPELWGG L VDLPADPESL
1201 DASALYAVLR GDGGEDQVAL RRGAVLGRRL VPDATPDVAP GSSPDVSGGA
1251 AHADATSGEW QPHGAVLVTG GVGHLADQVV RWLAASGAEH VVLLDTGPAN
1301 SRGPGRNDDL AAEEAEHGTE LTVLRSLSEL TDVSVRPIRT VIHTSLPGEL
1351 APLAEVTPDA LGA AVSAAAR LSELPGIGSV ETVLFFSSVT ASLGSREHGA
1401 YAAANAYLDA LAQRAGADAA SPRTVSVGWG IWDLPDDGDV ARGAAGLSRR
1451 QGLPPLEPQL ALGALRAALD GGKGHTLVAD IEWERFAPLF TLARPTRL LD
1501 GIPAAQRVLD ASSES AEASE NASALRRELT ALPVRERTGA LLDLVRKQVA
1551 AVLRYEPGQD VAPEKAFKDL GFDSL VVVEL RNRLRAATGL RLPATLVYDY
1601 PTPRTLAAHL LDRVLPD GGA AELPVA AHLD DLEAALTDLP ADDPRRKGLV
1651 RRLQTLLWKQ PDAMGAAGPA DEEEQAAPED LSTASADDMF ALIDREWGTR
1701 *

MonH, probable regulatory protein Length: 981 amino acids

1 VSGVERGVGS AGPVEQGDGL AGLVERAEAL AALRGAFDGS PGTGGSLVVL
 51 SGAVGTGKTA LLRAWADRIG ADADALVLTA TACRAERDLP LGVLEQLVRS
 101 PGLPPASAER ALAWWDEEAS ATPGKTDANG TSANGTDANG TGAGQTGAGQ
 151 AGVGQTGVGG EPVLAASALR GLCEVLRDLL AERPVVVAVD DAHHADAASL
 201 QCLLSVVRRL RSARLHVLFY EYAHQKAQNA LLSSEFLHEP ALRRIRLEPL
 251 SKAGVEALLA RHLDERTAQD LTPVVHGMSA GHPLLVRALA EDHRAAGGAG
 301 EAYGRAVLSF LYRHETPVTQ VARAIAALGA HAGPGQVGRL LDVDAASVER
 351 AVRQLTVAEV LHEGRLCHPA FAAAVLDGMP PEERRALHGR VADLLHEEGA
 401 PATEVAAHLV AADRSDAPWA VPFVQEAAQL ALDEDQVETG VDYLRAAHQR
 451 CRGAAQRAAV VGALADAEWR LDKAKVLRHL PDPAAMAPQT DPAALAPHTD
 501 PAPTAAPTAA PTPTPIPTTP PLPTHLLWHG RVEEGLDAIG TLTGPGPNPA
 551 GAPPMPNADL DTPWLWGAYL YPGHVKERLG SGALSPQRST PPAVTPELQG
 601 AGTLMNDLLH GGERDATEAA ERALNRYRLG PRTIAVQTAA LAALTYRDRP
 651 HRAAAWCDGL VAQADERNSP TWRALFTAWR ALLHLRQGDV AAAEQRAETA
 701 LALLGSKGWG AAIGLPLAAA VQAKAALGDV DGAAALLERP VPQAVFQTRT
 751 GLHYLAARGR YHLATGCHYA ALCDFYACGT RMSSWGVDLP ALEPWRLGAA
 801 EAYLALGEGE LARQLVDGQL PLPTPDDGRT WGMTLRLRAA TSPAPARAEL
 851 LDEAVAVLRE SGDTFELARA VADQAVAVRE GGEAERARLL ARKAELLARR
 901 WGSAPAPATV PEPPERPGPA TPDAELTSAE RRVAELAAEG FTNREISRKL
 951 CVTVSTVEQH LTRIYRKLDV RRLDLQAALG *

MonCI, flavin-dependent epoxidase Length: 496 amino acids

1 VTTTRPAHAV VLGASMAGTL AAHVLAHVLD AVTVVERDAL PEEPQHRKGV
 51 PQARHAHLLW SNGARLIEEM LPGTTDRLLA AGARRLGFPE DLVTLTGQGW
 101 QHRFPATQFA LVSARPLLDL TVRQQALGAD NITVRQRTEA VELTGSGGGS

151 GGRVTGVVVR DLDSGRQEQL EADLVIDATG RGSRLKQWLA ALGVPAALEED
 201 VVDAGVAYAT RLFKAPPGAT THFPAVNIAA DDRVREPGRF GVVYPIEGGR
 251 WLATLSCTRG AQLPTHEDEF IPFAENLNHP ILADLLRDAE PLTPVFGSRS
 301 GANRRLYPER LEQWPDGLLV IGDSLTA FNP IYGHGMSSAA RCATTIDREF
 351 ERSVQEGTGS ARAGTRALQK AIGA AVDDPW ILAATKDIDY VNCRVSATDP
 401 RLIGVDTEQR LRFAEAITAA SIRSPKASEI VTDVMSLNAP QAELGSNRFL
 451 MAMRADERLP ELTAPPFLPE ELAVVGLDAA TISPTPTPTP TAAVRS*

MonBII, carbon-carbon double bond isomerase Length: 141 amino acids

1 MPDEAARKQM AVDYAERINA GDIEGVLDLF TDDIVFEDPV GRPPMVGKDD
 51 LRRHLELAVS CGTHEVPDPP MTSMDDRFVV TPTTVTVQRP RPMTFRIVGI
 101 VELDEHGLGR RVQAFWGVTD VTMDDPAGPA DTHPEGIRA *

MonBI, carbon-carbon double bond isomerase Length: 144 amino acids

1 MNEFARKKRA LEHSRRINAG DLDAIIDLYA PDAVLEDPVG LPPVTGHDAL
 51 RAHYEPLLAH HLREEAAEPV AGQDATHALI QISSVMDYLP VGPLYAERGW
 101 LKAPDAPGTA RIHRTAMLV I RMDASGLIRH LKSYWGTS DL TVLG*

MonAVIII, polyketide synthase multienzyme MONS 8, housing extension modules 11 and 12 Length: 3754 amino acids

1 MSNEEKLLDH LKWVTAELRQ ARQRLHDKES TEPVAIVGMA CRYPGGARS A
 51 EDLWELVRDG GDAVAGFPDD RGWDLES LYH PDPEHPATSY VRDGAFLYDA
 101 GHFDAEFFGI SPREATAMDP QQRLLLETAW EAIEHAGMNP HALKGS DTGV
 151 FTGVS AH DY L TLISQTASDV EGYIGTGNLG SVVSGRISYT VGLEGP AVTV
 201 DTACSSSLVA IHLASQALRQ GECSLALAGG STVMATPGSF TEFSRQRGLA
 251 PDGRCKPFAA AADGTGWGEG AGVVALELLS EARRRGHKVL AVIRGSATNQ
 301 DGTSNGLAAP NGPSQERVIR AALANARLSA EDIDAVEAHG TGTTLGDP IE

351 AQALIATYGO GRPEDRPLWL GSVKSNIGHT QAAAGVAGVI KVMAMRNGL
401 LPTSLHIDAP SPHVQWEQGS VRLLEPVDW PAERTRRAGI SAFGISGTNA
451 HLILEEAPPE EDAPGPVAAE PGGVVPWVVS GRTPDALREQ ARRLGEFAAG
501 LADASVSEVG WSLATTRALF DQRAVVVGRD LAQAGASLEA LAAGEASADV
551 VAGVAGDVGP GPVLVFPQGQ SQWVGMGAQL LDESPVFAAR IAECEQALSA
601 HVDWSLSDVL RGDGSELSRV EVVQPVWAV MVSLAAVWAD YGITPAAVIG
651 HSQGEMAAAC VAGALSLEDA ARIVAVRSDA LRQLQGHGDM ASLSTGAEQA
701 AELIGDRPGV VVAAVNGPSS TVISGPPEHV AAVVADAEAQ GLRARVIDVR
751 YASHGPQIDQ LHDLLTDRLA DIQPTTTDVA FYSTVTAERL DDTTALDTAY
801 WVTNLRQPVR FADTIEALLA DGYRLFIEAS PHPVLNLGIQ ETIEQQAGAA
851 GTAVTIPTLR RDHGDTTQLT RAAAHAFATAG APVDWRRWFP ADPTPRTVDL
901 PTYAFQHKHY WVEPPAAVAA VGGGHDPVEA RVWQAIEDLD IDALAGSLEI
951 EGQAESVGAL ESALPVLSAW RRRHREQSTV DSWRYQVTWK HLPDVPAPEL
1001 SGAWLLLPA AHADHPAVLA TAQTLTAHGG EVRRHVVDAR AMERTELAQE
1051 LRVLMGAAF AGVVNLLALD EEPHPEHSAV PAGLAATTAL VQALADNGAD
1101 IAVRTLQGA VSTSAGDAL HPVQAQVWGL GRVAALEYPR LWGGLVDLPA
1151 RIDHQTLARL AAALVPQDED QISIRPSGVH ARRLAHAPAN TVGSGLGWRP
1201 DGTTLITGGT GGIGAVLARW LARAGAPHLL LTSRRGPDAP GAQELAAELT
1251 ELGAAVTVTA CDVGDREQVR RLIDDVPAEH PLTAVIHAAG VPNYIGLGDV
1301 SGAEDEVLR PKALAAHHLH ELTRELPISA FVMFSSGAGV WSGSQQAYG
1351 AANHFLDALA EHRRAEGLPA TSIWGPWAE AGMAADQAAL TFFSRFGLHP
1401 LSPELCVKAL QQALDAGETT LTVANFDWAQ FTSTFTAQRP SPLADLPEN
1451 RRASAPAAQQ EDATEASSLQ QELTEAKPAQ QRQLLLQHVR SQAAATLGHS
1501 DVDAVPATKP FQELGFDSL AVELRNRLNK STGLTLPTTV VFDHPTPDAL

1551 TDVLRaelSG DAAASADpVR AAGASrgaAD DEPIaIVGMA CRYPGDVRSA
1601 EELWDLVAAG KDAMGAfPDD RGWDLEtLYD PDPEsRGtSY VREGGfLYDA
1651 GDFDAGffGI SPREAVAMDP QQRLlLEtAW EAIERAGLDR ETLKGSDAGV
1701 FTGLtIFDYL ALVGEQpTEV EGYIGtGNLG CVASGRvSYV LGLEGPAMtI
1751 DTGCSSSLVA IHQAaHALRQ GECSLALAGG ATVMATPGsF VEFSLQRGLA
1801 KDGRCKPFAA AADGTGWAEG VGLVVLERLS EARRNGHNVL AVIRGSAINQ
1851 DGTSNGLTAP NGQAQQRVIR QALANARLSA EDVDAVEAHG TGTMLGDPIE
1901 ASALVATYgK ERPADRPLWL GSIKSNIGHA QASAGVAGVI KMVMALRNEQ
1951 LPASLHIDAP TPHVDWDGSG VRLlSEPvSW PRGERPRRAG VSAFGISGTN
2001 AHLILEQAPD APEPVTAPAE DAAAPAGVVP WVVSARGEeA LRAQARLLAD
2051 RATADPRLAS PLDVGWSLVK TRSVFENRAV VVGKDRQtLL AGLRSLAAGE
2101 PSPDVVEGAV QGASGAGPVL VFPGQGSQWV GMGAQLlDES PVFAARIAEC
2151 ERALSAHVDW SLSAVLRGDG SELSRVEVVQ PVLWAVMVSL ASVWADYGIT
2201 PAAVIGHsQG EMAAACVAGA LSLEDAARIV AVRSDALRQL MGQGDMAStG
2251 AGSEQVAELI GDRPGVCVAA VNGPSStVIS GPPEHVAAvV ADAEARGLRA
2301 RVIDVGYASH GPQIDQLHDL LTERLADIRP TTTDVAfYST VTAERLDDtT
2351 TLDTDYWVTN LRQPVRfADT IEALLADGYR LFIEASpHPV LNLGMEETIE
2401 RADMPATVVP TLRRDHGDAA QlTRAAaQAF GAGAEVDWtG WFPaVPLPRV
2451 VDLPTyAFQR ERfWLEGRRG LAGDPAGLGL ASAGHPLlGA AVELADGGSH
2501 LLTGRISPRD QAWLAeHRVM DTVLLPGsAF VELALQAaVR AGCAELAElt
2551 LHTPLAFGDE GAGAVDVQvV VGSVAEDGRR PVTVHSRPTG EGEEAVWtRH
2601 AAGVVAPPGP DAGDASfGGT WPPPGATpVG EQDPYGELAS YGYDFGPGsQ
2651 GLVSAWRLGD DLfAEVALPE AESGRADRYQ VHPVLLDATL HALILDAVtS
2701 SADTDQVLLP fSWSGLRVHA PGAEKLRVRI ARTAPDQLAL TAVDGGGGGGE

2751 PVLTL[•]ESL[•]TV RPVAAHQIAG ARAADRDALF RLVWMEVAAR AEETGGGAPR
 2801 AAVLAPVESG PMGGTSAGAL ADALSDALAA GPVWDTFGAL RDGVAAGGEA
 2851 PDVVLAVCAA PGAGAGAVAD ADGRGGDPAG YARLATVSLL SLLKEWVDDP
 2901 AFAATRLVVV TRGAVAARPG ETAGDLAGAS LWGLVRSQA ENPGRLTLLD
 2951 VDGLESSPAT LTGVLASGEP ELALRDGRAY VPRLVRDDAS VRLVPPVGS
 3001 TWRLARCQEA GGGQQLSLVD APEAGRALEP HEVRVAVRAA APGPLTAGQV
 3051 EGAGVVTEVG GEVGSVAVGD RVMGLFDAVG PVAVTDAALL MPVPAGWSWA
 3101 QAAGSLGAYV SAYHVLADV APRGGETLLV GEETGSVGRA VLRLALAGRW
 3151 RVEAVDGAST ADDSGAERAA DVTLRHEGAL VVHRAGGRPD EGQAVVPPEP
 3201 GRVREILAE TELTELA[•]EIT ESAEPGLPAE RGDSRALTPL DITVWDIRQA
 3251 PAAMAAPPSA GTTVFSLPPA FDPEGTVLVT GGTGALGSLT ARHLVERYGA
 3301 RHLLLSRRG ADAPGALELA ADLSALGARV TFAACDPGDR DEAAALLAAV
 3351 PSDHPLTAVF HCAGTVNDAV VQNLTAEQVE EVMRVKADAA WHLHELTRDA
 3401 DLSAFVLYSS VAGLLGGPGQ GSYTAANAFL DALARHRHDG GAAATSLAWG
 3451 YWELASGMSG RLTDADRARH ARAGVVGLGA DEGLALLDAA WAGGLPLYAP
 3501 VRLDLARMRR QAQSH PAPAL LRD[•]LVRGGSK SGGGAVSAGA AALLKSLGAM
 3551 SDPEREEALL DLVCTHIAAV LGYDAATPVN ATQGLRELGF DSLTAVELRN
 601 RLSAATGLKL PATFVFDHPN PAELAAQLRQ ELAPRAADPL ADVLAEFERI
 3651 EDSLLSVSSK DGSARAELAG RLRATLARLD APQDTAGEVA VATRTRIQDA
 3701 SADEIFAFID RDLGRDGASG QGNGQPTGQG NGHNGNGNG NGNGHGQAVE
 3751 GQR*

MonAVII, polyketide synthase multienzyme MONS 7, housing extension module 10 Length: 1642 amino acids

1 MAHTEEKLL EYLKRV TADLR QTERRLQDVE SAGHEPVAVI GMACRLPGGV
 51 RSPEEFWELV STGGDAVAPL PGNRNWDLDS LYDPDPESTG TSYVREGGFV

101 YDAGDFDPTF FGIGPTEAAA MAPQORLAL E TAWEAIERAG IDPLSLRSSD
151 TSTFIGCDGL DYALGASEVP EGTAGYFTIG NSGSVTSGRV AYTLGLEGPA
201 VTVDTACSSS LVSLHLATQA LRTQECSLAL AGGTYVMSSP APLIGFSELR
251 GLAPDGRCKP FSASSDGMGM AEGTGVVLE RLSDARRKGH KVLAVIRGSA
301 INQDGASNGL TAPNGPAQER VIRAAANAR LAPEDIDAVE AHGTGTTLGD
351 PIEAGALISA YGRERPEDRP LWVGAVKSN I GHTQIAAGVA GVIKMWLALR
401 HDLLPAILHV DAPSPHVEWD GSGLRLLTDP VKWPRGERPR RAGVSSFQFS
451 GTNAHLILEE APPEEEDVPG SVAEEP GGVP PWVVSGRTPD ALRAQARRLG
501 EFAAGPADAS AADV GWSLTT TRSVFEHRAV VVGRDRDALT AGLGALAAGE
551 ASAGVVAGVA GDVGPGPVLV FPGQGSQWVG MGAQLLDESP VFAARIAECE
601 RALSAYVDWS LSAVLRGDGS ELSRVEVVQP VLWAVMVSLA AVWADYGVTP
651 AAVIGHSQGE MAAACVAGAL SLEDAARIVA VRSDALRRLQ GHGDMASLST
701 GAEQAAELIG DRPGVVAAV NGPSSTVISG PPEHVA AVVA DAEARGLRAR
751 VIDVGYASHG PQIDQLHDLL TERLADIRPA NTDVAFYSTV TAERLTDTTA
801 LDTDYWVTNL RQPVRFADTI EALLADGYRL FIEASAH PVL GLGMEETIEQ
851 ADIPATVVPT LRRDHGDTTQ LTRAAAHAFT AGAPVDWRRW FPADPTPRTV
901 DLPTYAFQHQ HYWLERSASA SGA VSGEQSA AEAQLWHAVE ELDLGLLAET
951 LGSEEGSEEA VRALEPALPV LKGWRRRHQD QATIDSWRYR VTKQRS DGP
1001 APELGGDWLL FVPADKAEHP AVRATAEALS EHGA AAVRLH PVETGRAGRQ
1051 ELAAVDTAGL AGIVNLLALD EEPHPEHPAV PAGLAATTAL LQALGDNGTT
1101 APLHTVTQGA VSTGATDPLT HPLQAHVWGL GRVA ALEHPR LWAGLVDLPA
1151 RIDRHTLPRL AAALLPQDDE DQTAVRPTGI HHRRLTHAVG SIQNPVHSEA
1201 TWRPRGTTLI TGGTGGIGAV LARWLARQGA PRLHLTSRRG PDAPGARELA
1251 AELDGLGTAV TITACDVSDP RQLSGLIDDM PAEHPLTAVI HAAGMTDLTA

1301 IGDLTARLG EVLGSKSDAA WNLHELTRDL DLSAFVMFSS GAGVWGSGQQ
1351 GAYGAANHFL DALAEHRR AQ GLPATSI AWG PWAEAGMSAD PESLTYFKRF
1401 GLLPIAPDLC VKALHQAVDA GDATLTVANF DWAKFTPTFT AQRPSFLDD
1451 LPENQREAEQ TGTAETS AF REELAKTPAS QRLGFLVQQV RTYAAATLGR
1501 TVEDIPAAKP FQELGFDSL T AVQLRNQLNT TTGLSLPATV IFDHPTPEAL
1551 ATHLRGQLGD GAEVAGEGDV LAALDKWDTA FGAAEVDEAA RRRIVGRLQV
1601 LVSKWSPAQD GPEGTDSAHA DLEAASADDI FDLISSEFGK S*

MonD, cytochrome P450 hydroxylase Length: 431 amino acids

1 VGLTVGPDNA KRGIVPITDS KPAATFPDLV DPSFWARPHA ERVALFEEMR
51 GLPRPAFIRQ NMPGVPWTFG YHALVKYADI VEVSRRPQDF SSNGATTIIG
101 LPPELDEYYG SMINMDNPEH SRLRRIVSRS FGRNMIPEFE AVATRTARRI
151 IDELIARGPG DFIRPVAAEM PIAVLSDMMG IPAEDHDFLF DRSNTIVGPL
201 DPDYVPDRAD SERAVIEASR ELGDYIAGLR AERLAAPGND LITKLVQVQA
251 DGEQLTRQEL VSFFILLVIA GMETTRNAIS HALVLLTEHP EQKQLLLSDF
301 DTHAPNAVEE ILRVSTPINW MRRVATRD CD MNGHRFRRGD RIFLEYWSGN
351 RDESVPDPY RFDITRG TNA HVTFGAVGPH VCLGAHLARM EITVLYRELL
401 AALPQIHAVG QPRRLDSSFI EGIKHLHCAF *

MonRI, probable activator protein Length: 268 amino acids

1 VRYEMLGPLR IKDGNDYATI NAQKVEIVLT VLLIRADRVV SLEQLMREIW
51 GEDLPRRATA GLHVIISQLR KFLKVP GSAG NPVETRAPGY VLHKRDDDQI
101 DAQIFPELVD VGRSLLREKR FDEAASC FGQ ALALWRGPIL GQGGNGPGTN
151 GPIIDGFSTW LTEIRLECQE MLVE CQLQLG RHREAVGMLY ALTAENPMCE
201 AFYRQLMLAL YRSERQADAL KVYQSVR KTL NDELGLEPGR PLQELQRAIL
251 AGDMHLMSP PLALSGR*

MonAX, thioesterase Length: 278 amino acids

1 LSAFLAKGKI LSAFPPPDMS DPWIRRRFRPR PEAVVRLVCF PHAGGSASY
51 HPLAQSP TLP TDSEVLAVQY PGRQDRRRER LLDDIGELAD LITDALGPFD
101 DRPLAFFGHS MGAVLAYEVA QRLRERTGKQ PCRLFVSGRR APSRFRRGTV
151 HLLDDTELA ELRRAGGTDP RFLDDEELLA EIIPVVRNDY RAVELYRWNP
201 SPPLSCPITA LVGDRDPQAP LDEVEAWQQH TEGPFDLKVF AGGHFYLNTH
251 QQGVTEVISK ALADSAQQRA TARGNAR*

ORF29, a homologue of CapK involved in cell wall biosynthesis Length: 428 amino acids

1 LADLVAHARS ASPYYRELYH GLPERIEDPT LLPVTDKKQL MDHFDDWPTD
51 RDITFEKVRA FTDDPELIGR RFLGRYLVAT TSGTSGRRGL FVLDDRYMNV
101 SSAVSSRVLA SWLGPLGIAR AVVHGGRFAQ LVATEGHYVG FAGYSRLRQD
151 GEARSKLVRA FSVHEPMSRL VAELENEYRPA FVIGYASTIM LFTAEQEAGR
201 LHIDPVLVEP AGETMTESDT DRIAAAFGAK VRTMYSATEC TYLSHGCAEG
251 WYHVNDWAV LEPVDADHRP TPPGEFSHTT LISNLANRVQ PFLRYDLGDS
301 VMLRPDPCPC GTPSPAIRVQ GRSGDILTFP SGRGDDVSLA PLAFSSLFDR
351 MPGVELFQIE QTAPSTLRVR VVQAPGADAD HVWQRAHDGL THLLADNKLD
401 NVTVERGEEP PRQASGGKYR TIIPLAA*

LipB, lipase B Length: 338 amino acids

1 VKVPVEVTVR LSSWLGLVA AVLAATVLP SAASAADVSS PPLEIPAAEL
51 AKALHCGTEL GDLRDAGDKP TVLFVPGTGL KGEENYAWNY MAELKKKGYQ
101 SCWVDSPGRG LRDMQESVEY VVYATRAIQE ATGRKVDLVG HSQGGLLTAW
151 ALRFWPDLPG KVDDMVTLGS PFQGTRLASP CRPIAEVAGC PASVLQFARD
201 SNWSKALGAD GTPMPAGPSY TTIYSYADES VVADGEAPSL PGAHRIGVQD

251 ICPGRPWP TH IAMVVDQVS Y DLVADAIEHP GPADTSRIDR AHCAKPMPL
301 NSQEAVDALP GLLNFPIELL IHSQPWVDEE PPLRPYAR*

ORF31, putative ion pump Length: 309 amino acids

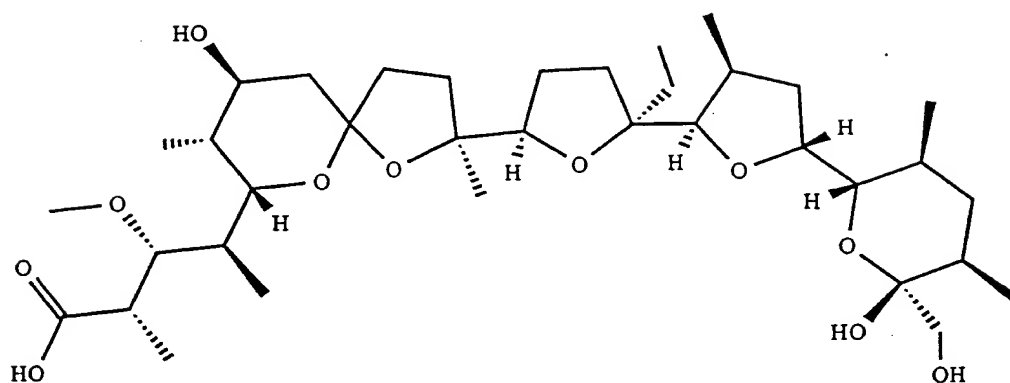
1 MGHDHGPSAG AAGGTLSGTY RKRLLTIGI SGSITVIQVV GALLSGSLAL
51 LADAAHSLTD AVGVSLALGA ITLAQRAPT RRTFGFCRVE IFSAVLNALL
101 LVVIFAWVLW SAIGRFSEPV EVKGGLMFVV ALGGLAANLV GLWLLRDAKE
151 KSLNLRGAYL EVLGDALGSV AVIVGGLVIL LTGWQAADPI ASIVIGLLIV
201 PRAYGLLRDS LHVLLLEATPQ DVDLGEVRRH LLEERG VVAV HDLHGWTVTS
251 GMPVLT AHVV VTEEALASGY GELLGRLQRC VGGHFDVAHS TIQLEPEGHV
301 EEDGALHT*

ORF32, hypothetical membrane protein Length: 364 amino acids

1 MTRALTLHDW IVAGIAVVAG VVAGLLLRAL LRWLGERASK TRWSGDDVIV
51 DALRTLVP CA AITAGLAAAA GALPLTPRTG RNVTMTLTAL LILAATLTAA
101 RIVTGLVKAV AQSRSGVAGS ATIFVNITRV VVLAMGFLIV LQTLGISIAP
151 LLTALGVGGL AVALALQDTL ANLFAGVHIL AAKTVQPGDY IQLSSGEEGY
201 VVDINWRNTT VRQLSNNLVI IPNAKLAGTN MTNYSRPEQE LSIMVQVGVS
251 YDSDLEQVEK VTTEVVDEVM AEITGAVPDH EAAIRFHTFG DSRISFTVIL
301 GVGEFSDQYR IKHEFIKRLH QRYRAEGIRV PAPVRTVRVQ QGELPPPLGI
351 PHQRTSTQA RLH*

**AmtA, glycine amidinotransferase (partial coding sequence)
Length: 131 amino acids**

1 MSPVNSHNEW DPLEEIIVGR LEGATIPSSH PVVACNIPTW AARLQGLAAG
51 FEYPQRLIEP AQQELDQFIA LLQSLDVTVR RPAAVDHKHR FGTPDWQSRG
101 FCNSCPRDSM LVVGDEIIET PMAWPCRCFE T



monensin A : R = ethyl
monensin B : R = methyl

Fig 1

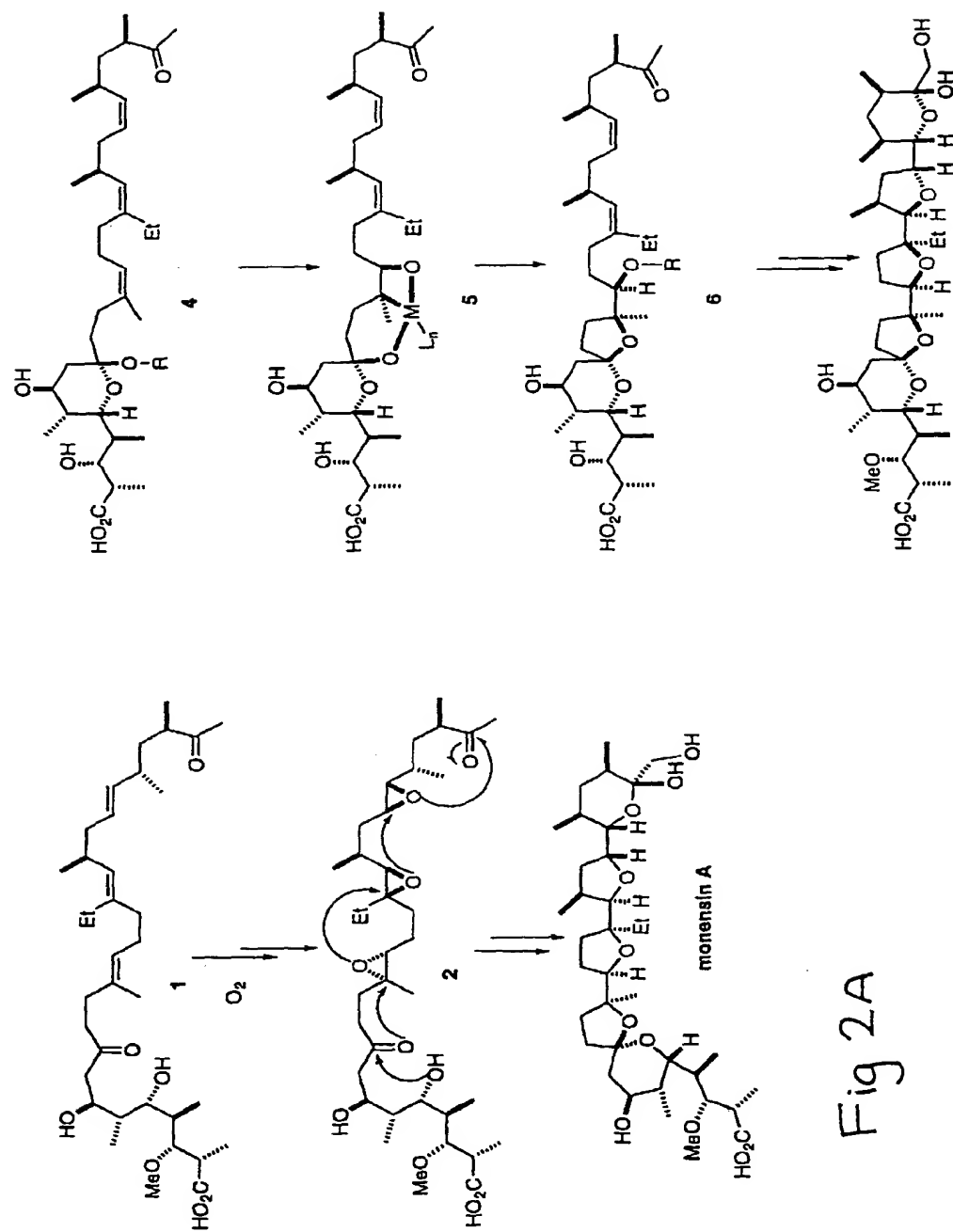


Fig 2A

Fig 2B

Figure 2. Proposed mechanisms for monensin biosynthesis.



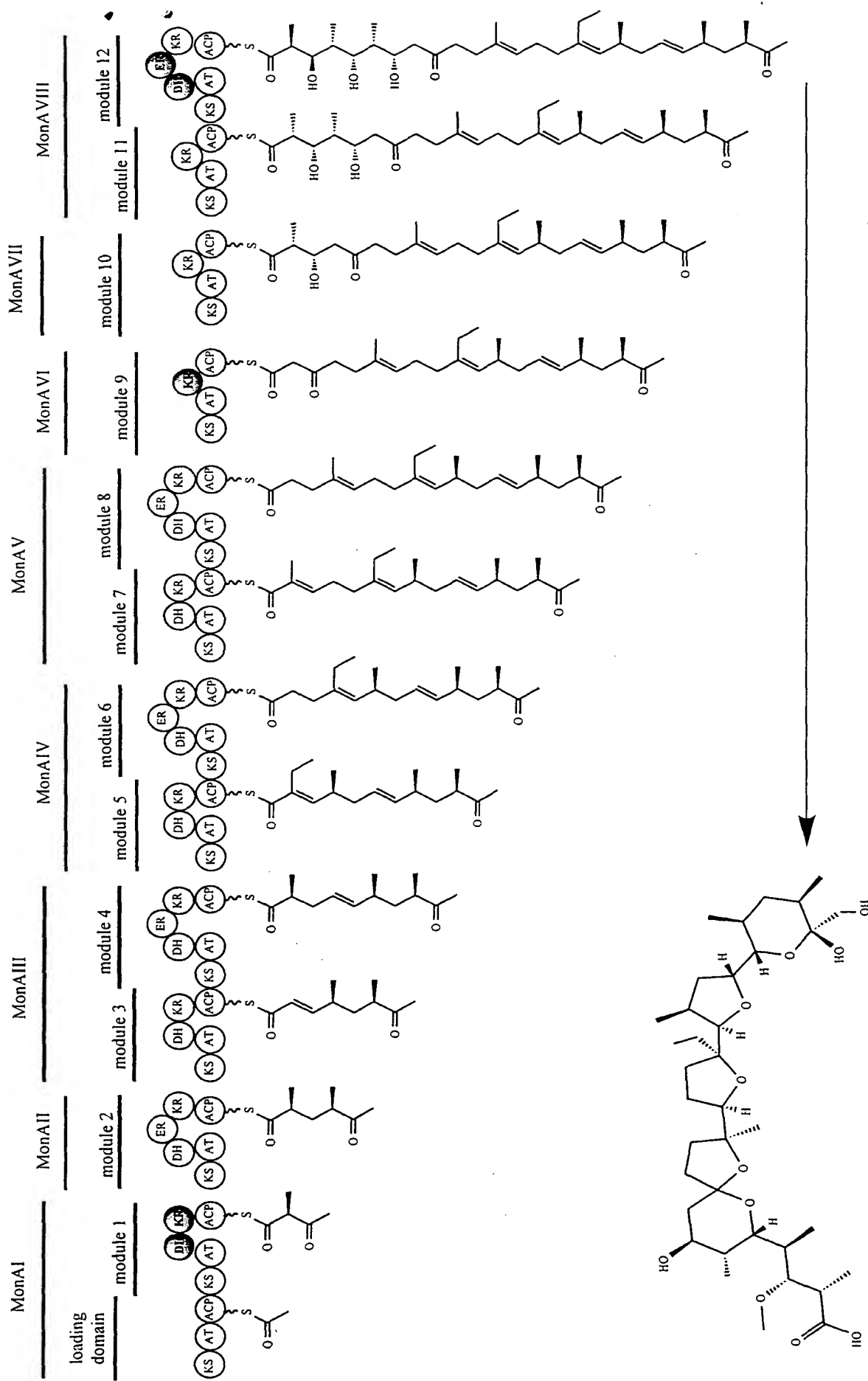


Fig3

Proposed organisation of the monensin PKS



Organisation of the Monensin Biosynthetic Gene Cluster

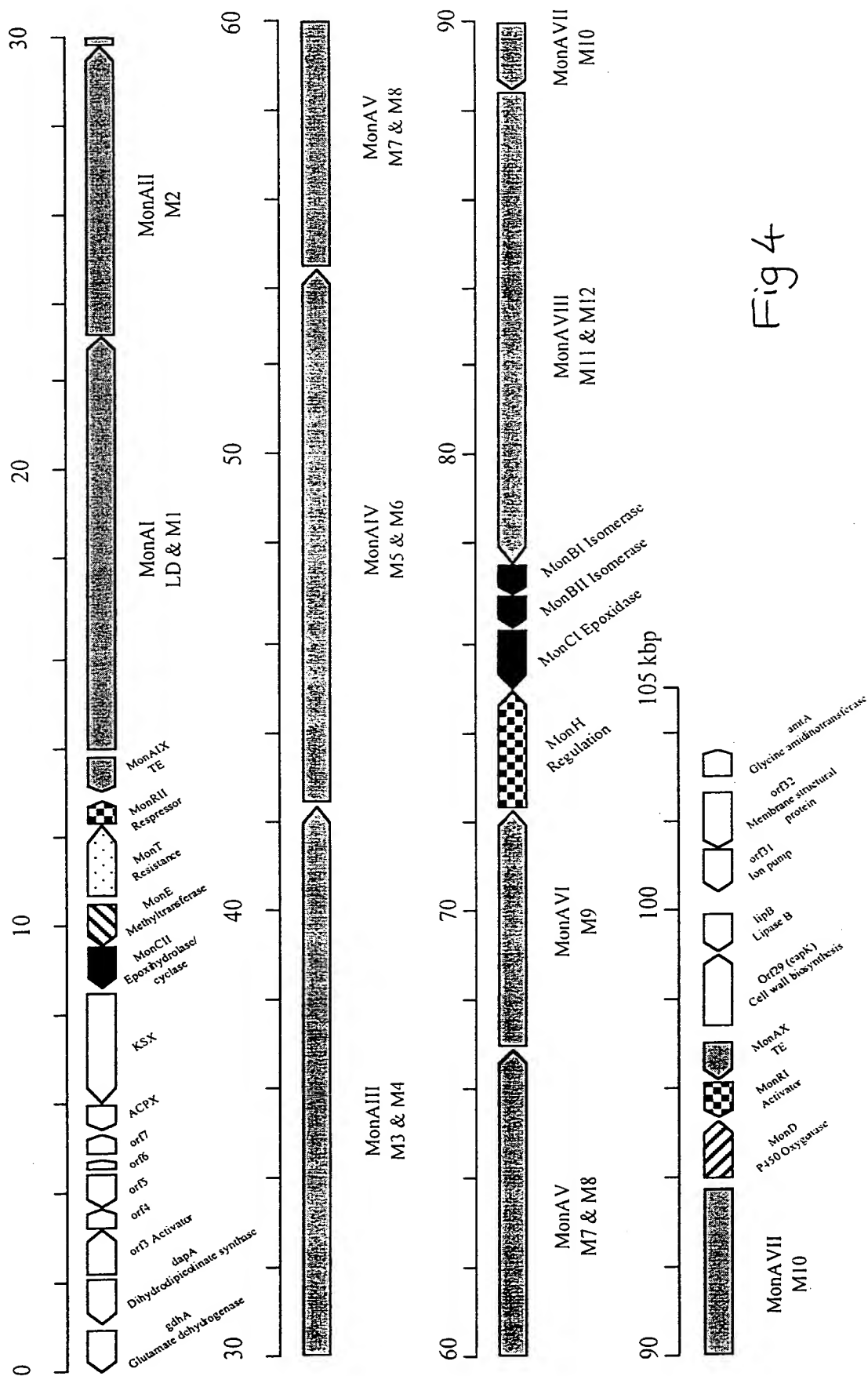


Fig 4

PCT/GB Not given

30/5/00

Mewburn Ellis.